

Result No.	Score	Query Match	Length	DB ID	Description
1	100	100	100	1	...
2	95	95	95	2	...
3	90	90	90	3	...
4	85	85	85	4	...
5	80	80	80	5	...
6	75	75	75	6	...
7	70	70	70	7	...
8	65	65	65	8	...
9	60	60	60	9	...
10	55	55	55	10	...
11	50	50	50	11	...
12	45	45	45	12	...
13	40	40	40	13	...
14	35	35	35	14	...
15	30	30	30	15	...
16	25	25	25	16	...
17	20	20	20	17	...
18	15	15	15	18	...
19	10	10	10	19	...
20	5	5	5	20	...

FEATURES		Location/Qualifiers	
source		1. .2821	
		/organism="Trichoplusia ni"	
		/db_xref="taxon:7111"	
CDS		26..2449	
		/codon_start=1	
		/product="insect intestinal mucin IIM22"	
		/db_xref="pid:g2224921"	
		/translation="MIKLLFLTLALGLVAAREVSDAEKNPALHEPHDPDCPAEQHWL LPHEVDCIKFYCEYGLKFIAPRDCAPGTEFKFSQATCVHAALAGCTILPGPAETTQA PATTAQPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQ PTTQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQ AATTAATPGVPAATSPVMPICELLNPGPADFDIHLIIPHKYCNLFYQCSNGY TFEORCPGLFNPVYQCDSPANVECDGEISPAVPVTEGNEDEDIDGLDNGCPA NFEIDLPHGNRCDKYOCVHGNLVERRCAGTHFSFELOQCDHIELVGCGLPGGES EEVDVDEACGHWYCPTEPIEWEPLNPGPADFSIDLHPHESDCGQYLQCVHGOTIA RPCGNLHFSNATQSCFSPVAGQVFECDSNQCTSTAAPTAAPTAAPTAAPTAAPT AAPTVVPATPAATAAPVPTTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT AAPTAVPEITVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT APNTVTPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT VAPAPNTAAPTVTTTSAPTATPEDDDIDPLPNPDPINPCVEECNVLPAWADCDKYW CDGNNOVLVVCSEGLQFNPTTKTCDPACNVGCVRSNIQMSSEYGVQVFIWNKLEID IRQALNFEI"	
BASE COUNT		710 a	1030 c 529 g 552 t
ORIGIN			
Query Match 100.0%; Score 2821; DB 3; Length 2821;			
Best Local Similarity 100.0%; Pred. No. 0;			
Matches 2821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	1	GAAGAATAACCAAGCAAGTTATGATAAGAGCCCTCTATTCTGACGGCCCTCGG	60
DB	1	GAAGAATAACCAAGCAAGTTATGATAAGAGCCCTCTATTCTGACGGCCCTCGG	60
QY	61	GCTGTCGGCGGCGTCTCTGAAGTACGACGCGGAGAGAACCCCGCTCTCCACGAGCC	120
DB	61	GCTGTCGGCGGCGTCTCTGAAGTACGACGCGGAGAGAACCCCGCTCTCCACGAGCC	120
QY	121	GCACCCAGACTGCCCTCCGCTGAGCAGCAGTGGCTCTGCTCAGCAATACGACTGCAC	180
DB	121	GCACCCAGACTGCCCTCCGCTGAGCAGCAGTGGCTCTGCTCAGCAATACGACTGCAC	180
QY	181	CAAGTTCTACTACTGTGAATATGGTCTCAAGTTTCATCGCACCGAGAGACTGTGCTCTGG	240
DB	181	CAAGTTCTACTACTGTGAATATGGTCTCAAGTTTCATCGCACCGAGAGACTGTGCTCTGG	240
QY	241	TACCGAATTCAGTTCTCCGCTCAGACTTGTGTTTCAGCGCGCTTTAGCCGGATGCACCT	300
DB	241	TACCGAATTCAGTTCTCCGCTCAGACTTGTGTTTCAGCGCGCTTTAGCCGGATGCACCT	300
QY	301	GCAGGACCTCCAGCTGAGACACCCAGGCGCCCAACCACTCAGGCGCCCAACACAC	360
DB	301	GCAGGACCTCCAGCTGAGACACCCAGGCGCCCAACCACTCAGGCGCCCAACACAC	360
QY	361	CCAGGCGCCCAACCACTACTCAGGCGCCCTACTACAACCAACCCAGCGCCCAACCAAC	420
DB	361	CCAGGCGCCCAACCACTACTCAGGCGCCCTACTACAACCAACCCAGCGCCCAACCAAC	420
QY	421	CACCCAGGCGCCCAACCACTACTCAGGCGCCCAACCACTACTCAGGCGCCCAACCACTCA	480
DB	421	CACCCAGGCGCCCAACCACTACTCAGGCGCCCAACCACTACTCAGGCGCCCAACCACTCA	480
QY	481	GGCCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC	540
DB	481	GGCCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTAC	540
QY	541	CCAGGCGCCCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	600
DB	541	CCAGGCGCCCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	600
QY	601	CCCAACTACAATCACCAGGCTGCACTACTCCCGCGCGCACTACTCCCGCGCGCACTAC	660



Qy	1741	TGTCCTGAAATCCCAATCACTGTACATCAGCGGCTACCGGTGCCCCACCCAGCTGCCCC	1800
Db	1741		1800
Qy	1801	TGTCCTGAAATCCCAATCACTGTACATCAGCGGCTACCGGTGCCCCACCCAGCTGCCCC	1860
Db	1801		1860
Qy	1861	CACGGCTGCCCCACCGCAGCCCTACTACTGCTGTCCAGAAATCCCAACTACTGTGCAC	1920
Db	1861		1920
Qy	1921	ATCACCACTTACTGTGCCCCCACTACCGCAGCACTGCCCCCAACACACACAGTGCACCT	1980
Db	1921		1980
Qy	1981	ATCACCACTTACTGTGCCCCCACTACCGCAGCACTGCCCCCAACACACACAGTGCACCT	2040
Db	1981		2040
Qy	2041	CACGTGTGCCCCCACTGCAGCTCCCCCTACCGCTCGCACTGCACCCACACACAGTGC	2100
Db	2041		2100
Qy	2101	CCCGGTAACTACAAACGAGCGCACAGTACCAACACCTGAAGATGATGACATCGACCCGCC	2160
Db	2101		2160
Qy	2161	TCGCCCAACGACCCCATCAACCCCTGGCTGTGAAGATGCAACGTTTGCCCTGGGCTCA	2220
Db	2161		2220
Qy	2221	CGGTGACTGCGCAAAATACTGGGTCTGTGACGGCAACCAAGTCCTGGTGGTTGCTC	2280
Db	2221		2280
Qy	2281	TGAGGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTCGCTTGCAACGTCGGTTG	2340
Db	2281		2340
Qy	2341	CGTGAGGAGCAACATTCAGATGCTGAAAGCTACCAAGGAGTCCAGGTTCTCATCCCATG	2400
Db	2341		2400
Qy	2401	GAACAACTAGATTGAAGACATCAGACAGCGCTGAACCTTTGAGTTGTAAACCTACTTTAA	2460
Db	2401		2460
Qy	2461	TTAATGAAGTTTGTGTTTATTTTGTAGTTATTTTCCAATGGCGGGGAAAGTCCGCCAT	2520
Db	2461		2520
Qy	2521	TATGGGCTTGCAGCTTTTGAGGAACCTTTTTTTTACTACCAACATCTCTGTGAACC	2580
Db	2521		2580
Qy	2581	CATATTTATTACGTATTAAACATCGGTATTGAAACAGTTACATGATTTTTTCATTAAAT	2640
Db	2581		2640
Qy	2641	TTGAACAATTTATGTTGTTTGTCTCATTTAAATATCAAAATCATTTTCGAAACCTGG	2700
Db	2641		2700
Qy	2701	CAATTTGGATTGGAATATACAAATCGGTTAAGAAAAAAGATTTCTTAAAAATGT	2760
Db	2701		2760
Qy	2761	ATTATATATAAATGTATAATAATATACAAATTTAGCATTTAAAAAATAAAAAAATAA	2820
Db	2761		2820

[illegible]

QY 241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTACGGCGCTTTAGCGGATGCAACCT 300  
DB TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTACGGCGCTTTAGCGGATGCAACCT 312  
QY 301 GCCAGGACCTCCAGCTGAGACAAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAAACAC 360  
DB GCCAGGACCTCCAGCTGAGACAAACCCAGGCCCCAGCAACAACTCAGGCCCCAACAAACAC 372  
QY 361 CCAGGCCCCAACCACTACTCAGGCGCTACTCAACACCACCCAGGCCCCAACCAACAC 420  
DB CCAGGCCCCAACCACTACTCAGGCGCTACTCAACACCACCCAGGCCCCAACCAACAC 432  
QY 421 CACCCAGGCCCCAACCACTCAGGCCCCAACCAACCCAGGCCCCAACCACTCAGGCCCCAAC 480  
DB CACCCAGGCCCCAACCACTCAGGCCCCAACCAACCCAGGCCCCAACCACTCAGGCCCCAAC 492  
QY 481 GGGCCCTACTACTACCACTCAGGCCCCAACCAACCACTCAGGCCCCAACCACTCAGGCCCCAAC 540  
DB GGGCCCTACTACTACCACTCAGGCCCCAACCAACCACTCAGGCCCCAACCACTCAGGCCCCAAC 552  
QY 541 CCAGGCCCCAACCACTCAGGCCCCAACCACTCAGGCCCCAACCACTCAGGCCCCAACCACTCAGG 600  
DB CCAGGCCCCAACCACTCAGGCCCCAACCACTCAGGCCCCAACCACTCAGGCCCCAACCACTCAGG 612  
QY 601 CCACTACTACCACTCAGGCTGCAACTACCCGGCGGCAACTACCCGGCGGCAACTAC 660  
DB CCACTACTACCACTCAGGCTGCAACTACCCGGCGGCAACTACCCGGCGGCAACTAC 672  
QY 661 CCGGGCGCACTACCCCTGCGCGGCAACACCGCGGTGCAACTACCCAGGTTGCTGCG 720  
DB CCGGGCGCACTACCCCTGCGCGGCAACACCGCGGTGCAACTACCCAGGTTGCTGCG 732  
QY 721 ACCCACTTACGCCCCAGTCTGCCCGCGGCTGTGAACTGTTGCCCAATGGTTGCCCGACG 780  
DB ACCCACTTACGCCCCAGTCTGCCCGCGGCTGTGAACTGTTGCCCAATGGTTGCCCGACG 792  
QY 781 TGACTTCGACATCCACTTGTGATTTCCCAACGACAACTGCAACTCTTCTTACCACTG 840  
DB TGACTTCGACATCCACTTGTGATTTCCCAACGACAACTGCAACTCTTCTTACCACTG 852  
QY 841 CTCACAGGTTACACTTCGACAGAGTGCCCTGAGGACTCTACTTCAACCCCTTACGT 900  
DB CTCACAGGTTACACTTCGACAGAGTGCCCTGAGGACTCTACTTCAACCCCTTACGT 912  
QY 901 CCAGGCTCGGACTCTCTCTGCTTAACTTGAATGCGAGCGGCAATCAGCCCGGCAACCC 960  
DB CCAGGCTCGGACTCTCTCTGCTTAACTTGAATGCGAGCGGCAATCAGCCCGGCAACCC 972  
QY 961 AGTCAGAGGCAACGAAGAGGAGACATTTGACATCGGAGACTCTCTCGACAATGGATG 1020  
DB AGTCAGAGGCAACGAAGAGGAGACATTTGACATCGGAGACTCTCTCGACAATGGATG 1032  
QY 1021 CCCAGCTTAATCGAATCGACTGGCTTGTGCCCCAGGAAACCGTTGCGAGCAAGTATTA 1080  
DB CCCAGCTTAATCGAATCGACTGGCTTGTGCCCCAGGAAACCGTTGCGAGCAAGTATTA 1092  
QY 1081 CCAGTCCGTCACGGTAACTTGTGAGAGGCGTGTGAGCGCGGCACTTCACTTCACTT 1140  
DB CCAGTCCGTCACGGTAACTTGTGAGAGGCGTGTGAGCGCGGCACTTCACTTCACTT 1152  
QY 1141 TGAATTCAGCAATGTGACCACTCGAGCTCGTGTGCTGCACTCTCCCGGGGGGAGAG 1200  
DB TGAATTCAGCAATGTGACCACTCGAGCTCGTGTGCTGCACTCTCCCGGGGGGAGAG 1212  
QY 1201 CGAAGAAGTTGAGCTGAGAGAGTCCCTGACCGGCTGGTACTGCCCGGCAAGACCCAT 1260  
DB CGAAGAAGTTGAGCTGAGAGAGTCCCTGACCGGCTGGTACTGCCCGGCAAGACCCAT 1272  
QY 1261 TGAATGGAGGCCCCCTCCCAAGGCTGCCCTCGGACTTTCAGACTGACCACTCTCTCC 1320  
DB TGAATGGAGGCCCCCTCCCAAGGCTGCCCTCGGACTTTCAGACTGACCACTCTCTCC 1332

QY 1321 CCACGAGAGCGACTGGGCCAGTATCTACAGTGTGTCCATGACAGACTATGCGAAGACC 1380  
DB CCACGAGAGCGACTGGGCCAGTATCTACAGTGTGTCCATGACAGACTATGCGAAGACC 1392  
QY 1381 TTGCCCTGGAAACCTGCACTTTCAGTCTGCGCCACACAGTCTCTGTGTGACCCG 1440  
DB TTGCCCTGGAAACCTGCACTTTCAGTCTGCGCCACACAGTCTCTGTGTGACCCG 1452  
QY 1441 TGGTTGCCAAAGTTTTCGAGTGGATTCTGACAAACAGTGCACATCGACTGCTGCCCGGAC 1500  
DB TGGTTGCCAAAGTTTTCGAGTGGATTCTGACAAACAGTGCACATCGACTGCTGCCCGGAC 1512  
QY 1501 AGTGTCTCCAAAGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCC 1560  
DB AGTGTCTCCAAAGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCC 1572  
QY 1561 CACCGTGTGCCCACTGCAACGCCACCGCAACTGAGGCCCGAGTCCCACTTACAACCGC 1620  
DB CACCGTGTGCCCACTGCAACGCCACCGCAACTGAGGCCCGAGTCCCACTTACAACCGC 1632  
QY 1621 AATTCCTACTCCGGCGCCACCGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCC 1680  
DB AATTCCTACTCCGGCGCCACCGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCC 1692  
QY 1681 ATCCCAACCACTGTACAGTACCACTACTGCTGCTCCACCGCAGCGCTTACTACTGCT 1740  
DB ATCCCAACCACTGTACAGTACCACTACTGCTGCTCCACCGCAGCGCTTACTACTGCT 1752  
QY 1741 TGTCCCTGAAATCCAAATCACTGTACATCAGCGCTACCGGTGCCCGCCACCGCTGCC 1800  
DB TGTCCCTGAAATCCAAATCACTGTACATCAGCGCTACCGGTGCCCGCCACCGCTGCC 1812  
QY 1801 CACCGCTGCCCGCCACCGGAGCGCTTACTGCTGTCCAGAAATCCCACTTACTGCTAC 1860  
DB CACCGCTGCCCGCCACCGGAGCGCTTACTGCTGTCCAGAAATCCCACTTACTGCTAC 1872  
QY 1861 ATCACCACTTACTGCTGCCCGCCACCTACCGCAGCACTGTGCCCGCAACCACTTACTGCT 1920  
DB ATCACCACTTACTGCTGCCCGCCACCTACCGCAGCACTGTGCCCGCAACCACTTACTGCT 1932  
QY 1921 ACCACCACTTGTGCTGCCCGCTTACTACCGCAGCACTGTGCCCGCAACCACTTACTGCTAC 1980  
DB ACCACCACTTGTGCTGCCCGCTTACTACCGCAGCACTGTGCCCGCAACCACTTACTGCTAC 1996  
QY 1981 ACCCAACCGAGCGCTTACTACCGCAGCACTGTGCCCGCAACCACTTACTGCTAC 2040  
DB ACCCAACCGAGCGCTTACTACCGCAGCACTGTGCCCGCAACCACTTACTGCTAC 2055  
QY 2041 CACTGCTGCCCGCTTACTGCTGCCCGCTTACTGCTGCCCGCAACCACTTACTGCTAC 2100  
DB CACTGCTGCCCGCTTACTGCTGCCCGCTTACTGCTGCCCGCAACCACTTACTGCTAC 2160  
QY 2101 CCGGTAACTTACAACCGGAGCGCACTTACCACTTGAAGATGATGATGATGATGATGATGATG 2160  
DB CCGGTAACTTACAACCGGAGCGCACTTACCACTTGAAGATGATGATGATGATGATGATGATG 2176  
QY 2176 CGTGACTGCGCAAAATACTGGGTCTGTGAGCGCAACCACTTACTGCTGCTGCTGCTGCTGCTG 2280  
DB CGTGACTGCGCAAAATACTGGGTCTGTGAGCGCAACCACTTACTGCTGCTGCTGCTGCTGCTG 2295  
QY 2281 TGAGGCTTCCAGTTTCAACCGGCTTACTTAAAGCTGTGACTGCTGCTGCTGCTGCTGCTGCTG 2340  
DB TGAGGCTTCCAGTTTCAACCGGCTTACTTAAAGCTGTGACTGCTGCTGCTGCTGCTGCTGCTG 2355  
QY 2341 GGTGAGGAGCACTTCAAGATGCTGAAAGCTTACCAAGGCTTCAAGGCTTCAATCCCATG 2400  
DB GGTGAGGAGCACTTCAAGATGCTGAAAGCTTACCAAGGCTTCAAGGCTTCAATCCCATG 2460  
QY 2401 GAACAACTAGATGAAGACATCAGACAGGCGCTGAACTTTGAGTTGTAAACCTTACTTAA 2460







Qy	364	GGCCCCAACACAACTACTCAGGCCCTTACTATAACCAACCCAGGCCCCCAACACCAACACAC 423
Db	5748	GACCCCAACCCCAACACCCACCGGCACACAGACCCCAACACGACGACCATCAGCACCCAC 5807
Qy	424	CCAGGCCCCCAACCAACCAACCCAGGCCCCCAACCAACCAACCAAGGCCCCCACTACCCTCAGGC 483
Db	5808	CACTACGGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 5867
Qy	484	CCCTACTACTACCACTCAGGCCCAACCAACCAACCACTCAGGCCCTTACCAACCAACCAACCA 543
Db	5868	CACCAACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 5927
Qy	544	GGCCCCAACCAACCAACCCAGGCCCCCAACTACCAACCCAGGCCCACTACCACTCAGGC GCCC 603
Db	5928	GACACCCATCAACACC--ACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 5984
Qy	604	AACTACAACTACCCAGCTGCAACTACCCCGCGCGGAAGTACCCCGCGCGGAAGTACCCCGCGGAAGTACCCCG 663
Db	5985	GACCCCAACCAACCAACCCATCACCACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCA 6044
Qy	664	GGCGCAACTACCCCTGCGCGCAACCCCGCTGCACTACCCCAAGTGTTCCTGTGACC 723
Db	6045	CGGCACACAGACCCCAACCAACCAACCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6104
Qy	724	CACCTTCAACCCCACTGTGGCCCCCGCATCTGTGAAGTGTTCGCCAATGGTTGCCACG---- 780
Db	6105	AACACCAACCGGCACACAGACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6164
Qy	780	-CTGACTTCGACATCACTGTGTGATTCCCAACGACAAGTACTGCAACTCTTCTACCAAG 838
Db	6165	CCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6224
Qy	839	TGCTCCAAACGGTTACACCTTCGAACAGAGGTGCCCTGAGGAGCTTACTTTCAACCCCTAC 898
Db	6225	TACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6281
Qy	899	GTCCAGCGCTCGCACTCTCTGTGTAAGTTGAAGCGGCGGAATCAGCCCCCGCACCC 958
Db	6282	CACCAACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6341
Qy	959	CCAGTCAAGAAGCAACGAAGACGAAAGATTCATGATGGAGACCTCTCTGCACAATGGA 1018
Db	6342	GACACCCATCACCACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6401
Qy	1019	TGCCCACTAACTTCGAATCGACTGGCTCTTGCCCAACGGAACCGTTGGCAAGTAT 1078
Db	6402	CCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6461
Qy	1079	TACCAGTCGCTCCACGGTAACCTGGTGAAGAGGGGTTGTGAGCGCGGACCCCACTTCAGT 1138
Db	6462	CACACAGACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6521
Qy	1139	TTTGAATTCAGCAATGTGACCACTCGAGCTCGTTGGCTGCACCCCTCCCGGGGGGAG 1198
Db	6522	ACCCACCGGCACACAGACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6581
Qy	1199	AGCGAAGAAGTTGAGTCGACGAGGATGCTTGACCGGGTGTACTGCCCCACGGAACCC 1258
Db	6582	AACCCCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCA 6641
Qy	1259	ATTGAATGGAGGCCCTCCCAACGCTGCCCTGCGGACTTCAGATCGACCACTCTCTC 1318
Db	6642	GGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6701
Qy	1319	CCCCACGAGAGCACTCGGCCCACTATCTACAGTGTGTTCATGGACAGACTATCGCAAGA 1378
Db	6702	CACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6759
Qy	1379	CCTTGCCTGGAAACCTGCACTTCACTCTCGGCACACAGTCTCTGTGAGTCTCTGTGACC 1438
Db	6760	CCCATCACCAACCACTACGGTGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCCCA 6819

QY	1439	GCTGGTTCGCAAGTTTTCGAGTGCAGATCTTGACAAACCAAGTGCACATCGACTGCTGCGCCCG	1498
Db	6820	ACCACGACAGCCATTCACCAACCACTACGGGTGACCCCAACCCCAACCAACCAACCGGCACA	6879
QY	1499	ACAGCTGCTCAAGGGTGGCCCAACGGCTGCCCAAGGGTGGCCCAACTGCGCGCACCC	1558
Db	6880	CAGACCCCAACACGACGACCCATTCACCAACCACTACGGTGACCCCAACCCCAACACCC	6939
QY	1559	TCCACCTGGTTCACCTGCAACGGCCACCGCAACTGACGGCCC-----AGTCCCACT	1612
Db	6940	ACCGGCACACAGACCCCAACACGACACCCATTCACCAACCACTACGGTGACCCCAACCC	6999
QY	1613	ACAACCGCAATTCCTACTCGGGCCCCACCGTGGCCGCCCAACCGAGCTCTTACTACTGCT	1672
Db	7000	CCAACACCCACGGCAGACAGACCCCAACACGACACCCATCACCACCACTACGGTG	7059
QY	1673	GCCCCGAAATCCCAACCACTGTACAGTACACACTACTGTGTCCTCCACGCGAGCCCT	1732
Db	7060	ACCCCAACCCCAACACCCACCGGCACACAGACCCCAACACGACACCCATCACCACACC	7119
QY	1733	ACTACTGTGCTCCCTGAATCCCAATCACTGTGCATCATAGCGCTTACCGTGCGCCCAACC	1792
Db	7120	ACTAGGTGACCCCAACCCCAACACCCACCGGCACACAGACCCCAACACGACACCCATC	7179
QY	1793	GCTGCCCCCAACCGGTG---CCCCACCGCAGCCCTTACTTGTGTCGCCAAGATCCCA	1849
Db	7180	ACCACCACTACGGTGTACCCCAACCCCAACACCCACCGGCACACAGACCCCAACCAACG	7239
QY	1850	ACTACTGTACATCACACACTACTGTGCGCCGCACTACGGGAGACACTTGCGCCCAACACC	1909
Db	7240	ACAGCCATCACCAACCACTAGGGTGACCCCAACCCCAACACCCACCGGCACACAGACC	7299
QY	1910	ACAGTCACTGTACACCACTGTGCGCCCTAC---TACGGAGACCTGCTCCCAACACC	1966
Db	7300	CCACACGACACCACTACACCAACCACTACGGGTGACCCCAACCCCAACACCCACCGGC	7359
QY	1967	ACAGTGACTGCACCAACCCACCGCAGCCCTTACTTACCGAGACCTTGCGCCCAACACCA	2036
Db	7360	ACACAGACCCCAACCAACCACTACCAACCACTACGGTGACCCCAACCCCAACCAACCA	7419
QY	2027	GTCAGTGTACACCACTGTGCGCCCACTTGAGCTTCCCTTACGGTGCACCTGCACCC	2086
Db	7420	CCACCGGCACACAGACCCCAACACGAGACCACTACCAACCACTACGGTGACCCCA	7479
QY	2087	AACACCACTGCCCGGTAACTAACACCGGCACGACCTTACCACTTACCACTGAAGATGAT	2146
Db	7480	ACCCCAACCCACCGGCACACAGACCCCAACCAACGACACCCATCACCACCACTACG	7539
QY	2147	GACATCGACCCCTCTTCCCAACGACCCCAACCC	2184
Db	7540	GTGACCCCAACCCCAACCCACCGGCACACAGACCC	7577
RESULT	5		
AC002042/c			
LOCUS	AC002042	185994 bp	DNA
DEFINITION	*** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 16p11.2 clone C1978K-A-180G2; HTGS phase 1, 5 unordered pieces.		
ACCESSION	AC002042		
NID	g3075381		
KEYWORDS	HTG; HTGS_PHASE1.		
SOURCE	human.		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
AUTHORS	1 (bases 1 to 185994)		
TITLE	Fuhrmann, J., Kim, U. J., Kerlavage, A. R., and Venter, J. C.		
JOURNAL	Human chromosome 16p13 BAC clone C1978K-A-180G2		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 185994)		
TITLE	Adams, M. D. and Loftus, B. J.		
	Direct Submission		



JOURNAL Submitted (29-APR-1997) The Institute for Genomic Research, 9712  
Medical Center Dr., Rockville, MD 20850, USA  
COMMENT On Apr 22, 1998 this sequence version replaced gi:3068565.  
\*\*\* WARNING: Phase 1 High Throughput Genome Sequence \*\*\*  
\*\*\* This sequence is unfinished. It consists of 5 contigs for  
\* which the order is not known; their order in this record is  
\* arbitrary. In some cases, the exact lengths of the gaps  
\* between the contigs are also unknown; these gaps are presented  
\* as runs of N as a convenience only. When sequencing is complete,  
\* the sequence data presented in this record will be replaced  
\* by a single finished sequence with the same accession number.  
\* 1 2198: contig of 2198 bp in length  
\* 2199 2248: gap of unknown length  
\* 2249 46288: contig of 44040 bp in length  
\* 46289 46338: gap of unknown length  
\* 46339 110919: contig of 64581 bp in length  
\* 110920 110963: gap of unknown length  
\* 110970 145225: contig of 34256 bp in length  
\* 145226 145275: gap of unknown length  
\* 145276 185994: contig of 40719 bp in length.  
\* 185994 185994: contig of 40719 bp in length.  
FEATURES  
source  
1. 185994  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/clone="A-18062"  
BASE COUNT 51275 a 43334 c 43047 g 48137 t 201 others  
ORIGIN

Query Match 5.9%; Score 166.4; DB 18; Length 185994;  
Best Local Similarity 55.6%; Pred. No. 7.7e-19;  
Matches 367; Conservative 0; Mismatches 281; Indels 12; Gaps 2;  
Qy 1485 CGACTGCTGCCCGACAGCTGCTCCACGGCTGCCCAACGGCTGCCCAACGGCTGCC 1544  
Db 112122 CTACTCTCTGCGACAGCTGCTGCAACTGCGGATGCTGCTACTCTGCGCACAGCTACTG 112063  
Qy 1545 CAACTGCGGACACCTCCACCGTGTCCCACTGCAAGCGCCACCGCACTGCAAGCGCCAG 1604  
Db 112062 CAACTGCTGATGCTGCTAC-----TCCTCTGCCACAGCTACTGCAACTGCTGATGCTG 112009  
Qy 1605 TCCCACTACAAACGCAATTCCTACTCGGGCCCCACGGTGGCCCCACCGAGCTCCTA 1664  
Db 112008 CTACTCTCTGCGACAGCTATTGCAACTGCTGCTACTCTCTGCGCACAGCTACTGCAA 111949  
Qy 1665 CTACTGCTGCCCTGAATCCCGACACCTGTCACAGTACCACTACTGCTGCTCCACCG 1724  
Db 111948 CTGCTGCTACTCTCTCTGCGCACAGCTACTGCAACTGCTGCTGCTACTCTCTGCGCA 111889  
Qy 1725 CAGCCCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1784  
Db 111888 CAGTACTGCAACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111829  
Qy 1785 CCCCCAGGCTGCCCGCCACCGCTGCCCGCCACCGAGCCCTACTACTGCTGCTGCCAGAAA 1844  
Db 111828 CTACTCTCTGCGCACAGCTACTGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 111769  
Qy 1845 TCCCAAC-----TACTGTACATFACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1898  
Db 111768 CTGCAACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 111709  
Qy 1899 CCCCCAACACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1958  
Db 111708 CTCTCTGCGCACAGCTACTGCAACTGCGGATGCTGCTACTCTCTGCGCACAGCTACTGCAA 111649  
Qy 1959 CCAACACCACTGACTGCAACCCACCGAGCCCTACTACTACCGAGCACTGCGCCCA 2018  
Db 111648 CTGCTGATGCTGCTACTCTCTCTGCGCACAGCTACTGCAACTGCTGCTGCTGCTGCTGCT 111589  
Qy 2019 ACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2078

Db 111588 CTGCCACAGCTACTGCAACTGCTGATGCTGCTACTCTCTGCGCACAGCTACTGCAACTG 111529  
Qy 2079 CTGCCACCAACACACAGCTGCCCGGTAACATAACAGCGACCACTACCACTG 2138  
Db 111528 CTGATGCTGCTACTCTCTCTGCGCACAGCTACTGCAACTGCTGATGCTGCTACTCTCTCTG 111469

RESULT 6  
125041/c 125041 10596 bp DNA PAT 21-AUG-1996  
LOCUS Sequence 15 from patent US 5547856.  
DEFINITION 125041  
ACCESSION g1604911  
NID  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unknown.  
REFERENCE 1 (bases 1 to 10596)  
AUTHORS Godowski, P.J., Lokker, N.A. and Mark, M.R.  
TITLE Hepatocyte growth factor variants  
JOURNAL Patent: US 5547856-A 15 20-AUG-1996;  
FEATURES Location/Qualifiers  
source 1. 10596  
/organism="unknown"  
BASE COUNT 2627 a 2571 c 3023 g 2375 t  
ORIGIN

Query Match 5.6%; Score 158.8; DB 6; Length 10596;  
Best Local Similarity 53.3%; Pred. No. 1.8e-17;  
Matches 359; Conservative 0; Mismatches 312; Indels 3; Gaps 1;  
Qy 1496 CCGACAGCTGCTCCACAGGGTGGCCCAACGGCTGCCCAACGGCTGCCCAACGGTGGCCGCA 1555  
Db 2946 CCTCCACCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2887  
Qy 1556 CCTTCACAGTGTGCCACCTGCAACGGCCACCGCAACTGACAGCCCGAGTCCCACTACACA 1615  
Db 2886 GCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2827  
Qy 1616 ACGCCAATTCCTACTGCGGGCCCCACCGCTGCCCGCCACCGAGCTCTCTACTGCTGCTGCC 1675  
Db 2826 CTTGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2767  
Qy 1676 ---CCTGAATCCCCAACCACTGTGCACAGTACCACTACTGCTGCTGCCACGCGAGCCCT 1732  
Db 2766 CTTCTGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2707  
Qy 1733 ACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1792  
Db 2706 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2647  
Qy 1793 GCTGCCCGCCAGCTGCTGCCCGCCACCGAGCCCTACTACTGCTGCTGCCAGAAATCCCAACT 1852  
Db 2646 CTTGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2587  
Qy 1853 ACTGTCACATCACCACCACTACTGCTGCTGCCCGCCACTACCGCAGACCTGCCCGCCACACA 1912  
Db 2586 CTTGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2527  
Qy 1913 GTCACTGTACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1972  
Db 2526 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2467  
Qy 1973 ACTGCACACCGCAGCGCCCTACTACCGCAGACCTGCTGCCCGCCACACACACAGTCACT 2032  
Db 2466 CTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2407  
Qy 2033 GTACACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2092  
Db 2406 GCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2347  
Qy 2093 ACAGTGCCCGGTAACATAACAGCGCGCACAGCTACCACTGATGATGATGATGATGATGATGAT 2152









AUTHORS	Sequin,C., Farrell,P.J. and Barrell,B.G.	features or because of some other experimental data. The reading
TITLE	DNA sequence and transcription of the BamHI fragment B region of	frames are named according to the Bam HI fragment in which they
JOURNAL	B95-8 Epstein-Barr virus	start. eg BALF3 is the third leftward frame starting in Bam HI
MEDLINE	Mol. Biol. Med. 1 (3), 369-392 (1983)	fragment A. BORF1 is the first rightward frame in Bam HI fragment
REFERENCE	85060424	O. If there is an obvious TATA sequence followed by an in frame Met
AUTHORS	10 (bases 45415 to 52824)	codon that satisfies the rules of Kozak [12] in that there is a
TITLE	Jones,M.D., Foster,L., Sheedy,T. and Griffin,B.E.	purine at -3 and/or a G at +4 then the reading frame is numbered
JOURNAL	The EB virus genome in Daudi Burkitt's lymphoma cells has a	from the A of the ATG to the base preceding the termination codon.
MEDLINE	deletion similar to that observed in a non-transforming strain	If there is no obvious initiation codon or there is a substantial
REFERENCE	(P3HR-1) of the virus	reading frame in phase before the ATG then the reading frame is
AUTHORS	EMBO J. 3 (4), 813-821 (1984)	numbered from the first base of the first codon.
TITLE	84207939	
JOURNAL	11 (bases 87650 to 92703)	
MEDLINE	Biggin,M., Farrell,P.J. and Barrell,B.G.	SITES of POLYA signals
REFERENCE	Transcription and DNA sequence of the BamHI L fragment of B95-8	This feature lists all occurrences of the sequence AATAAA which is
AUTHORS	Epstein-Barr virus	found normally approximately 20 bases upstream of the mRNA
TITLE	EMBO J. 3 (5), 1083-1090 (1984)	processing/polyA addition site. The rarely used homolog AATAAA is
JOURNAL	84236104	only listed when it is found in a position close to the end of a
MEDLINE	12 (bases 7315 to 9312)	major reading frame.
REFERENCE	Yates,J., Warren,N., Reisman,D. and Sugden,B.	
AUTHORS	A cis-acting element from the Epstein-Barr viral genome that	SITES of DONOR and ACCEPT sequences
TITLE	permits stable replication of recombinant plasmids in latently	This is not a comprehensive listing of all such sequences and only
JOURNAL	infected cells	the positions of a few have been noted because they occur in
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)	potentially interesting positions. The number quoted in the table
REFERENCE	84222045	is the position of the terminal base in the intron in each case.
AUTHORS	13 (bases 76089 to 79808)	
TITLE	Gibson,T., Stockwell,P., Ginsburg,M. and Barrell,B.	Restriction enzyme SITES.
JOURNAL	Homology between two EBV early genes and HSV ribonucleotide	Only the positions of the sites Bam HI (BAM) are listed.
MEDLINE	reductase and 38K genes	
REFERENCE	Nucleic Acids Res. 12 (12), 5087-5099 (1984)	RPT
AUTHORS	84247360	This feature is used to define repetitive sequences.
TITLE	14 (bases 1 to 172281)	
JOURNAL	Baer,R., Bankier,A.T., Biggin,M.D., Deininger,P.L., Fallell,P.J.,	SITE DEL
MEDLINE	Gibson,T.J., Hatfull,G., Hudson,G.S., Satchwell,S.C., Seguin,C.,	This feature defines deletions in B95-8 with respect to other
REFERENCE	Tuffnell,P.S. and Barrell,B.G.	strains such as RAJI and also to deletions in other strains such as
AUTHORS	DNA sequence and expression of the B95-8 Epstein-Barr virus genome	P3HR1 and DAUDI with respect to B95-8.
TITLE	Nature 310 (5974), 207-211 (1984)	
JOURNAL	15 (bases 1 to 172281)	SITE HPN
MEDLINE	Bodscot,M. and Perricaudet,M.	Denotes sequences with twofold symmetry ie could form hairpin
REFERENCE	Clustered alternative splice sites in Epstein-Barr virus RNAs	loops. This is not a comprehensive list - only a few occurrences
AUTHORS	Nucleic Acids Res. 15 (14), 5887 (1987)	noted.
TITLE	87289053	ORGRPL
JOURNAL	16 (bases 1 to 172281)	Denotes the region that encompasses an origin of replication (ori
MEDLINE	Laux,G., Perricaudet,M. and Farrell,P.J.	P).[13].
REFERENCE	A spliced Epstein-Barr virus gene expressed in immortalized	
AUTHORS	lymphocytes is created by circularization of the linear viral	NUMBERING
TITLE	genome	The DNA sequence of B95-8 EBV has been revised [19]. The original
JOURNAL	EMBO J. 7 (3), 769-774 (1988)	(Baer et al, 1984) base 359 has been deleted so the new sequence
MEDLINE	88283646	around that position reads TCAGTCTTT. To avoid renumbering the
REFERENCE	17 (bases 1 to 172281)	entire sequence, position 1 has been moved 1 base to the left of
AUTHORS	Hatfull,G.F., Barrell,B.G., Quinn,J. and McGeoch,D.	the EcoRI site separating EcoRI Dhet from EcoRI I
JOURNAL	Unpublished	(ie the first A of AGAATTC).
REFERENCE	18 (bases 1 to 172281)	Location/Qualifiers
AUTHORS	Farrell,P.J. and Barrell,B.G.	1. .172281
TITLE	Direct Submission	/Organism="Human herpesvirus 4"
JOURNAL	Submitted (05-JUN-1984) to the EMBL/GenBank/DBJ databases	/Strain="B95-8"
MEDLINE	19 (bases 1 to 172281)	/db_xref="taxon:10376"
REFERENCE	Farrell,P.J.	
AUTHORS	Direct Submission	
TITLE	Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer	
JOURNAL	Research, St. Mary's Hospital Medical School, Norfolk Place London	
COMMENT	W2 1PG	
	CDS	
	Listed under this feature are all known protein coding regions as	
	well as all the major open reading frames in the sequence. In	
	general the term major is taken as the longest frame in a	
	particular region taking into account the adjacent longest frames	
	and likely transcription signals. Note that on this basis some long	
	overlapping frames have been excluded and on the other hand some	
	small frames have been included which might represent exons or	
	genes because they occur in a logical combination with other	

```

mRNA      58. .272
/note="exon 2 terminal protein RNA"
mRNA      360. .458
/note="exon 3 terminal protein RNA"
misc_feature complement(535)
mRNA      540. .788
/note="polyA signal: AATAAA"
mRNA      871. .951
/note="exon 4 terminal protein RNA"
mRNA      1026. .1196
/note="exon 5 terminal protein RNA"
promoter  complement(1192)
/note="TATA: TATAAAT"
mRNA      1280. .1495
/note="exon 7 terminal protein RNA"
promoter  complement(1383)
/note="TATA: CATAAAA"
mRNA      1574. .1682
/note="exon 8 terminal protein RNA"
promoter  1676
/note="TATA: TATAAAG"
promoter  1691
/note="TATA: TATATAA BN-R1 late promoter before BNRf1,
gives 4.1kb late RNA. Probably encodes non glycosylated
140kd protein in membrane antigen. Also two latent RNAs
spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson
et al, 1985). The longer one encodes terminal protein."
CDS       1736. .5692
/note="BNRf1 reading frame, 5 NKT/s"
/codon_start=1
/db_xref="PID:g59075"
/db_xref="SWISS-PROT:P03179"
/translation="MEERGREITQMPVARYGGFFIMVRLFGQDGEANIQERLYELLSD
PRSAAGDPGLIAELLVLRGTNDPQORARELALVGLLNGEGGHELGST
ESALASNNVYAYGPDMARPTWSAEIQOFLLGATYVLRVMGRQGFVEHRS
RPSFOQAIHLVFLDNALRYKDSQVAAGFORALLVAGPETADTRDPLKLEWVF
GGRAAGQOLADELKI VSALRDTYSCHLVLOPTELTDTWKVLSRDRTRAHSEHGFIH
AAGTIOANCOPLMEROHGFLPFVYNALASSLGWYQATGPGADARAAARQQAQFQ
RAACCHAKSGVPVAVGYRTINATLKGEGLOPTFMGELGAIKQALDVRIDYGH
VIEHLGSPQWGLPLPCCPYAESWAQAQVOTALFSLYPAPCISGYARPPGSA
FITRQGEKINGRTVLQALGRACDMAGCOHYVLGSLTVPLGLNFVNDLSPVSTAE
MDDFSPTFTVQTEQAGASPVLVDSDSDISPSYELPWLSECLTSLSHPTV
GSKHLVRHTDVSGRVAGQGVCLDPLADYAFVAHSQVWTPPGGAPLPLVETWD
RMTEKLLVSAFGGKRVKYSVTIYLGEQYKVSLLDREGTRLMAEALLNACAPIL
DPEDVLLTLHLDRADNAGVMEATAAADYARGLVKLTFFGASCPETGSSASF
MTVASVAPGEFSGLTIPVLQKTGSLLIARVCGDKRIQGGSLFEQLFSDVATTPRA
PEALSLKHLFRAVQQLKSGVILSGHDSIDGSLVCLVEMALAGQGVITIMPVASYD
LPMEAPHGLVFEVEERSVGEVLOTLSMNNYPVLVGRVGEQDQDFVQHGPEVTV
LROSLRLLLGWSSFASQYECLELRDRINRSMHVSIDYNEALAVSPLTKNLSRRL
VTEPPRCQAVLCAPTRGHESLLAFTNAGCLRCRRVFFREVRDNTFLDKYVGLAIG
GVGHADALSAGRAVIALINRFPALRDAILKFLNRPDFFSVAGLGVQVILAGLVAGV
STDNPPAPGVENVQORSLIILAPNAGSMGFESKWLNISIPATTSYMLRGLRGVLCPCW
VQSGCLGLQPTVQLNAHQIACHFHSNGTDARFAMNYPNRPTEQGNIGLCS
RDGRHLALCDPLCTDFQWEHIIPPAFGHGTGCSPTWLMFQAHLWSLRHGRPSE"
complement(1795)
/note="polyA signal: AATAAA"
misc_feature 3955
/note="BAM: Bam H1 Nhet/h"
misc_feature 3994
/note="BAM: Bam H1 h/C"
mRNA      5408. .5856
/note="exon 9 terminal protein RNA"
misc_feature 5841
/note="polyA signal: AATAAA, end of 4.1kb late RNA and TP
latent RNA."
misc_feature 5863
/note="alternative end to TP cDNAs"
promoter  6097
/note="TATA: TATAAGA"
misc_feature 6629. .6795
/note="Pol III RNA EBER 1"
promoter  complement(6823)
/note="TATA: CATAAAT"
misc_feature 6956. .7128
/note="Pol III RNA EBER 2"
rep_origin 7315. .9312
/note="origin of replication, ori P (Yates et al, 1984,
1985)"
repeat_region 7421. .8042
/note="21x30bp repeats, binding sites for EBNA-1 (site I,
Rawlins et al, 1985). Tandem repeat part of orlp (Reisman
et al, 1985). Also functions as a cell type specific
enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
7738
/note="TATA: TATAAAT"
7888
/note="TATA: TATAAAT"
8573
/note="TATA: TATAAAT"
misc_feature complement(8680)
/note="polyA signal: AATAAA"
misc_feature complement(8755)
/note="polyA signal: AATAAA"
8962
/note="polyA signal: AATAAA"
9021. .9133
/note="HPN: dyad symmetry, site II for EBNA-1 binding
(Rawlins et al, 1985). Dyad symmetry part of orlp (Reisman
et al, 1985)"
promoter  complement(9398)
/note="TATA: TATAAAT"
9631
/note="TATA: TATAAAT BC-R1 late promoter before BCRf1"
9675. .10187
/note="BCRf1 reading frame"
/codon_start=1
/db_xref="PID:g59076"
/db_xref="SWISS-PROT:P03180"
/translation="MEERLVTVLQCLVLYLAPECGTDCDNEPOMLRDLRDAFSRV
KTFQTKDEVNDLLKESLLEDGFGYLGCOALSEMIOFYLEEVMPQAEQNDEAKDHV
NSLGENLTKLRLLRRCHRFPCENKSKAVEQIKNAFNKLOEKGIYKAMSEFDIFIN
IEATYMTKAR"
10076
/note="TATA: GATAAAA"
complement(10148)
/note="polyA signal: AATAAA"
10173
/note="polyA signal: AATAAA"
10257
/note="polyA signal: AATAAA, end of 0.8kb late RNA from
BCR1 and end of 1.6 kb late RNA, start unknown"
complement(10277)
/note="polyA signal: AATAAA"
complement(10975)
/note="TATA: CATAAAT"
11305
/note="TATA: TACAAA; BCR2 promoter for highly spliced
EBNA latent RNAs."
11336. .11480
/note="exon C1 of Bodescot et al (1986) RNAs"
11524
/note="TATA: TATAAAT"
complement(11587)
/note="polyA signal: AATAAA"
complement(11606)
/note="TATA: CATAAAT"
11626. .11657
/note="exon C2 of Bodescot et al (1986) RNAs"
11796
/note="TATA: TATAAGT"
complement(11799)
/note="TATA: TATAAAA"
12001. .15072
/note="3072 repeat 1"

```

CDS  
12541..13689  
/note="ECRF2 3072 repeat, reading frame 1; Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25077"  
/db\_xref="PID:g1334836"  
/db\_xref="SPTREMBL:Q04397"  
/translation="VWEAEGRRPGEVEGDRGLCWQSPGDLRPSGGRSPSPAPQTD  
PRVSRQGPASSGAAGSPQAPQTRVSRADRPRAWLLGASRRGWCFCPSLCPEEPG  
TSGTPEPLGASRRRPPGLRSLSPVKKECLRGATLGAQAPESRGQGLRVPVRVPGQ  
PEGPRQGRPORVPVRPPGLQSPGCPPEGLTVSPPLQARASPSRRGASLGPQVOP  
HRDSGDPPTGSLCPAPLQPSLHPRPQLLASGPPGQEGPRQGRVAFPLPWL  
LPASHPSLSLPHRHVQAGRRDPGGVSPVPPAAQSLPPCKGASFSPSPURPSLLCT  
VCKVQPTTPVHGSRAQRPPLPTVDPRSVHPGHRPPVPSTPVPVSRGDFM"  
13215  
misc\_feature  
14352  
promoter  
/note="TATA: TATAAG BWR1 one of the promoters for highly spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al, 1986)"  
14384..14410  
/note="\*exon W0 of EBNA/LP RNAs"  
14554..14619  
/note="\*exon W1 (also W61) of EBNA/LP RNAs forms initiator met when fused to exon W0 or exon C2."  
14701..14832  
/note="\*exon W2 (also W132) part of LP gene"  
15073..18144  
/note="3072 repeat 2"  
15613..16761  
/note="BWR1 reading frame 2; Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25078"  
/db\_xref="PID:g1334837"  
/db\_xref="SPTREMBL:Q04397"  
/translation="VWEAEGRRPGEVEGDRGLCWQSPGDLRPSGGRSPSPAPQTD  
PRVSRQGPASSGAAGSPQAPQTRVSRADRPRAWLLGASRRGWCFCPSLCPEEPG  
TSGTPEPLGASRRRPPGLRSLSPVKKECLRGATLGAQAPESRGQGLRVPVRVPGQ  
PEGPRQGRPORVPVRPPGLQSPGCPPEGLTVSPPLQARASPSRRGASLGPQVOP  
HRDSGDPPTGSLCPAPLQPSLHPRPQLLASGPPGQEGPRQGRVAFPLPWL  
LPASHPSLSLPHRHVQAGRRDPGGVSPVPPAAQSLPPCKGASFSPSPURPSLLCT  
VCKVQPTTPVHGSRAQRPPLPTVDPRSVHPGHRPPVPSTPVPVSRGDFM"  
16287  
misc\_feature  
17424  
promoter  
/note="TATA: TATAAG"  
17626..17691  
/note="\*Exon W1"  
17773..17904  
/note="\*Exon W2"  
18145..18216  
/note="3072 repeat 3"  
18685..19833  
/note="BWR1 reading frame 3; Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25079"  
/db\_xref="PID:g1334838"  
/db\_xref="SPTREMBL:Q04397"  
/translation="VWEAEGRRPGEVEGDRGLCWQSPGDLRPSGGRSPSPAPQTD  
PRVSRQGPASSGAAGSPQAPQTRVSRADRPRAWLLGASRRGWCFCPSLCPEEPG  
TSGTPEPLGASRRRPPGLRSLSPVKKECLRGATLGAQAPESRGQGLRVPVRVPGQ  
PEGPRQGRPORVPVRPPGLQSPGCPPEGLTVSPPLQARASPSRRGASLGPQVOP  
HRDSGDPPTGSLCPAPLQPSLHPRPQLLASGPPGQEGPRQGRVAFPLPWL  
LPASHPSLSLPHRHVQAGRRDPGGVSPVPPAAQSLPPCKGASFSPSPURPSLLCT  
VCKVQPTTPVHGSRAQRPPLPTVDPRSVHPGHRPPVPSTPVPVSRGDFM"  
19359  
misc\_feature  
/note="BAM: BamH1 W/W"

promoter  
20496  
/note="TATA: TATAAG"  
20698..20763  
/note="\*Exon W1"  
20845..20976  
/note="\*Exon W2"  
21217..21288  
/note="3072 repeat 4"  
21757..22905  
/note="BWR1 reading frame 4; Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25080"  
/db\_xref="PID:g1334839"  
/db\_xref="SPTREMBL:Q04397"  
/translation="VWEAEGRRPGEVEGDRGLCWQSPGDLRPSGGRSPSPAPQTD  
PRVSRQGPASSGAAGSPQAPQTRVSRADRPRAWLLGASRRGWCFCPSLCPEEPG  
TSGTPEPLGASRRRPPGLRSLSPVKKECLRGATLGAQAPESRGQGLRVPVRVPGQ  
PEGPRQGRPORVPVRPPGLQSPGCPPEGLTVSPPLQARASPSRRGASLGPQVOP  
HRDSGDPPTGSLCPAPLQPSLHPRPQLLASGPPGQEGPRQGRVAFPLPWL  
LPASHPSLSLPHRHVQAGRRDPGGVSPVPPAAQSLPPCKGASFSPSPURPSLLCT  
VCKVQPTTPVHGSRAQRPPLPTVDPRSVHPGHRPPVPSTPVPVSRGDFM"  
22431  
misc\_feature  
23568  
promoter  
/note="TATA: TATAAG"  
23771..23835  
/note="\*Exon W1"  
23917..24048  
/note="\*Exon W2"  
24289..27360  
/note="3072 repeat 5"  
24829..25977  
/note="BWR1 reading frame 5; Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25081"  
/db\_xref="PID:g1334840"  
/db\_xref="SPTREMBL:Q04397"  
/translation="VWEAEGRRPGEVEGDRGLCWQSPGDLRPSGGRSPSPAPQTD  
PRVSRQGPASSGAAGSPQAPQTRVSRADRPRAWLLGASRRGWCFCPSLCPEEPG  
TSGTPEPLGASRRRPPGLRSLSPVKKECLRGATLGAQAPESRGQGLRVPVRVPGQ  
PEGPRQGRPORVPVRPPGLQSPGCPPEGLTVSPPLQARASPSRRGASLGPQVOP  
HRDSGDPPTGSLCPAPLQPSLHPRPQLLASGPPGQEGPRQGRVAFPLPWL  
LPASHPSLSLPHRHVQAGRRDPGGVSPVPPAAQSLPPCKGASFSPSPURPSLLCT  
VCKVQPTTPVHGSRAQRPPLPTVDPRSVHPGHRPPVPSTPVPVSRGDFM"  
25503  
misc\_feature  
26640  
promoter  
/note="TATA: TATAAG"  
26842..26907  
/note="\*Exon W1"  
26989..27120  
/note="\*Exon W2"  
27361..30432  
/note="3072 repeat 6"  
27901..29049  
/note="BWR1 reading frame 6; Protein sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25066"  
/db\_xref="PID:g1334841"  
/db\_xref="SPTREMBL:Q04397"  
/translation="VWEAEGRRPGEVEGDRGLCWQSPGDLRPSGGRSPSPAPQTD  
PRVSRQGPASSGAAGSPQAPQTRVSRADRPRAWLLGASRRGWCFCPSLCPEEPG  
TSGTPEPLGASRRRPPGLRSLSPVKKECLRGATLGAQAPESRGQGLRVPVRVPGQ  
PEGPRQGRPORVPVRPPGLQSPGCPPEGLTVSPPLQARASPSRRGASLGPQVOP  
HRDSGDPPTGSLCPAPLQPSLHPRPQLLASGPPGQEGPRQGRVAFPLPWL  
LPASHPSLSLPHRHVQAGRRDPGGVSPVPPAAQSLPPCKGASFSPSPURPSLLCT  
VCKVQPTTPVHGSRAQRPPLPTVDPRSVHPGHRPPVPSTPVPVSRGDFM"  
28575  
misc\_feature  
/note="BAM: BamH1 W/W"  
29712  
promoter

```
/note="TATA: TATAAAG"
29914..29979
/note="Exon W1"
30061..30192
/note="Exon W2"
30433..33504
/note="3072 repeat 7"
30973..32121
/note="BWRFl reading frame 7: Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25067"
/db_xref="PID:g1334842"
/db_xref="SPTREMBL:O04397"
/translat="VWEAGRRPRGCEVGDRLGCLWCQSPGDPRLRSPGGRSPSPAPQTD
PRVSRQGPASSGAASPPQAPQTRVVSASRADRPRAWLLGASRRGWFCPSLCSPSEPG
TSGTEPLGASRRPGLRSLPVSVPKRECLRGATLGAQAPESRGGHLRVPPRVPGQ
PEGPRQGPORPVRPPFGLQSPGCPPEGTGLVPSPLQARASPRRGSGLGQVQVP
HRDPSGDPPTGSLCPAPLQPSLHPRQLLASPGPGQEGPQGRGVAFPLPWPL
LPASHPSPLSLPPIHVRVHOAGRRDGGPVSPVAAAQSLPPGKGASFSPSLRPSLLCT
VCKVQPTPTVHGSRAQRPPLPTVDRPSVHGHPRPPVSTPVPSRGDFM"
misc_feature
31647..31647 "BAM: BamH1 W/W"
promoter
32784
/note="TATA: TATAAAG"
32986..33051
/note="Exon W1"
33133..33264
/note="Exon W2"
33505..33576
/note="3072 repeat 8"
34045..35193
/note="BWRFl reading frame 8: Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25068"
/db_xref="PID:g1334843"
/db_xref="SPTREMBL:O04397"
/translat="VWEAGRRPRGCEVGDRLGCLWCQSPGDPRLRSPGGRSPSPAPQTD
PRVSRQGPASSGAASPPQAPQTRVVSASRADRPRAWLLGASRRGWFCPSLCSPSEPG
TSGTEPLGASRRPGLRSLPVSVPKRECLRGATLGAQAPESRGGHLRVPPRVPGQ
PEGPRQGPORPVRPPFGLQSPGCPPEGTGLVPSPLQARASPRRGSGLGQVQVP
HRDPSGDPPTGSLCPAPLQPSLHPRQLLASPGPGQEGPQGRGVAFPLPWPL
LPASHPSPLSLPPIHVRVHOAGRRDGGPVSPVAAAQSLPPGKGASFSPSLRPSLLCT
VCKVQPTPTVHGSRAQRPPLPTVDRPSVHGHPRPPVSTPVPSRGDFM"
misc_feature
34719..34719 "BAM: BamH1 W/W"
promoter
35856
/note="TATA: TATAAAG"
36058..36123
/note="Exon W1"
36205..36336
/note="Exon W2"
36577..39648
/note="3072 repeat 9"
37117..38265
/note="BWRFl reading frame 9: Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25069"
/db_xref="PID:g1334844"
/db_xref="SPTREMBL:O04397"
/translat="VWEAGRRPRGCEVGDRLGCLWCQSPGDPRLRSPGGRSPSPAPQTD
PRVSRQGPASSGAASPPQAPQTRVVSASRADRPRAWLLGASRRGWFCPSLCSPSEPG
TSGTEPLGASRRPGLRSLPVSVPKRECLRGATLGAQAPESRGGHLRVPPRVPGQ
PEGPRQGPORPVRPPFGLQSPGCPPEGTGLVPSPLQARASPRRGSGLGQVQVP
HRDPSGDPPTGSLCPAPLQPSLHPRQLLASPGPGQEGPQGRGVAFPLPWPL
LPASHPSPLSLPPIHVRVHOAGRRDGGPVSPVAAAQSLPPGKGASFSPSLRPSLLCT
VCKVQPTPTVHGSRAQRPPLPTVDRPSVHGHPRPPVSTPVPSRGDFM"
37791
/note="BAM: BamH1 W/W"
38928
/note="TATA: TATAAAG"
39130..39195
/note="Exon W1"
39277..39408
/note="Exon W2"
39649..42720
/note="3072 repeat 10"
40189..41337
/note="BWRFl reading frame 10: Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25070"
/db_xref="PID:g1334845"
/db_xref="SPTREMBL:O04397"
/translat="VWEAGRRPRGCEVGDRLGCLWCQSPGDPRLRSPGGRSPSPAPQTD
PRVSRQGPASSGAASPPQAPQTRVVSASRADRPRAWLLGASRRGWFCPSLCSPSEPG
TSGTEPLGASRRPGLRSLPVSVPKRECLRGATLGAQAPESRGGHLRVPPRVPGQ
PEGPRQGPORPVRPPFGLQSPGCPPEGTGLVPSPLQARASPRRGSGLGQVQVP
HRDPSGDPPTGSLCPAPLQPSLHPRQLLASPGPGQEGPQGRGVAFPLPWPL
LPASHPSPLSLPPIHVRVHOAGRRDGGPVSPVAAAQSLPPGKGASFSPSLRPSLLCT
VCKVQPTPTVHGSRAQRPPLPTVDRPSVHGHPRPPVSTPVPSRGDFM"
40863
/note="BAM: BamH1 W/W"
42000
/note="TATA: TATAAAG"
42202..42267
/note="Exon W1"
42349..42480
/note="Exon W2"
42721..45792
/note="3072 repeat 11"
43261..44409
/note="BWRFl reading frame 11: Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25071"
/db_xref="PID:g1334846"
/db_xref="SPTREMBL:O04397"
/translat="VWEAGRRPRGCEVGDRLGCLWCQSPGDPRLRSPGGRSPSPAPQTD
PRVSRQGPASSGAASPPQAPQTRVVSASRADRPRAWLLGASRRGWFCPSLCSPSEPG
TSGTEPLGASRRPGLRSLPVSVPKRECLRGATLGAQAPESRGGHLRVPPRVPGQ
PEGPRQGPORPVRPPFGLQSPGCPPEGTGLVPSPLQARASPRRGSGLGQVQVP
HRDPSGDPPTGSLCPAPLQPSLHPRQLLASPGPGQEGPQGRGVAFPLPWPL
LPASHPSPLSLPPIHVRVHOAGRRDGGPVSPVAAAQSLPPGKGASFSPSLRPSLLCT
VCKVQPTPTVHGSRAQRPPLPTVDRPSVHGHPRPPVSTPVPSRGDFM"
43935
/note="BAM: BamH1 W/W"
45072
/note="TATA: TATAAAG"
45274..45339
/note="Exon W1"
45415..52824
/note="DEL: DAUDI deletion (Jones et al, 1984)"
45421..45552
/note="Exon W2"
45644..52450
/note="DEL: P3HR1 deletion (Jeang and Hayward, 1983)"
45793..47643
/note="3072 repeat 12"
46333..47481
/note="BWRFl reading frame 12: Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25072"
/db_xref="PID:g1334847"
/db_xref="SPTREMBL:O04397"
/translat="VWEAGRRPRGCEVGDRLGCLWCQSPGDPRLRSPGGRSPSPAPQTD
PRVSRQGPASSGAASPPQAPQTRVVSASRADRPRAWLLGASRRGWFCPSLCSPSEPG
TSGTEPLGASRRPGLRSLPVSVPKRECLRGATLGAQAPESRGGHLRVPPRVPGQ
PEGPRQGPORPVRPPFGLQSPGCPPEGTGLVPSPLQARASPRRGSGLGQVQVP
HRDPSGDPPTGSLCPAPLQPSLHPRQLLASPGPGQEGPQGRGVAFPLPWPL
LPASHPSPLSLPPIHVRVHOAGRRDGGPVSPVAAAQSLPPGKGASFSPSLRPSLLCT
VCKVQPTPTVHGSRAQRPPLPTVDRPSVHGHPRPPVSTPVPSRGDFM"
47007
misc_feature
```



```

/note="BAM: BamH1 W/Y"
47761..47793
/note="Exon Y1 Bodescot et al, 1984"
47831
/note="TATA: TATAAGT"
47878..47999
/note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and
Strominger,1985), last common exon"
complement(48023)
/note="polyA signal: AATAAA"
48386..48444
/note="exon Bodescot et al, 1984"
48386..50032
/note="Coding exon for EBNA-2 (Sample et al,1986),"
/codon_start=1
/db_xref="PID:e25073"
48504..49967
/note="BFRF1, encodes EBNA-2 (Dambaugh et al, 1984;
Dillner et al, 1984)"
/codon_start=1
/db_xref="PID:e25074"
/db_xref="PID:g1632787"
/db_xref="SWISS-PROT:P12978"
/translation="MPTFYLAHGGQTHLYVDTSLGNSLVSIPSNPYQELSDTP
LRDAMTEPSLDRDPLGYDVGHGLASAMRLMANYIVRQSGDRGLILPQGGQTAP
QARLVQHPVLPDRPTAPLILSPQSRLTPQPLMMPRPPTPTPLPATLVVPRPT
RPTTLPTLQVLPTELQPTSPPRMHLPLVLPDQSMHPLTHOSTPNDPDSPEP
RSPVTEYNIPMLPPSOLPPAAPQPPGVINDQQLHLLPSGPPWPPICDPPQPS
KIQSGRQSGRGGRGGRGKGRQKQRPQGPWRPEPNTSPSPSPSPSPVGLGH
QOQAGSDPTPGPSNAAPVCRNSHTATNPSPIHESHPNSPEAPILFPDDWYPPSID
PADLSDWDYIFETTESPDSDEYVEGPKRPRPSIQ"
48678..48800
/note="14 x CCCCCACCA repeats"
48848
/note="BAM: BamH1 Y/H"
49350
/note="TATA: TATAACA"
complement(49353)
/note="TATA: TATAAAA"
49525..49578
/note="9 x GGGGCA repeats"
49525..50032
/note="exon (Bodescot et al 1984)"
50003
/note="polyA signal: AATAAA, end of Bodescot T1 RNA and
EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"
complement(50156)
/note="TATA: TATAAGT"
complement(50317)
/note="polyA signal: AATAAA, end of 2.5kb early RNA from
52817"
complement(50578..52557)
/note="BHLFI early reading frame"
50578..52115
/note="12 x 125bp repeats"
52654..53697
/note="region homologous to Eco R1 C of Raji"
complement(52817)
/note="TATA: GATAAAA promoter for 2.5kb early RNA
containing BHLFI (Jeang and Hayward, 1983; Freese et
al,1983)"
53759
/note="TATA: TATAAC likely promoter for class III and IV
early RNAs encoding BHRF1 (Pearson et al, 1987)"
53895
/note="DONOR: CCGGTAAC donor for splice to 54335 in class
IV early RNAs encoding BHRF1 (Pearson et al, 1987)"
54335
/note="ACCEPT: TTWCTAG acceptor from 48444 in class I,
47999 in class II, and 53895 in class IV early RNAs
encoding BHRF1 (Pearson et al, 1987)"
54376..54948
/note="BHRE1 reading frame, limited homology to bcl-2 gene.
Early gene in B95-8 cells and part of restricted EA
complex."
54591
/note="TATA: TATAACA"
complement(54594)
/note="TATA: TATAAAT"
54853
/note="BAM: BamH1 H/F"
complement(54929)
/note="polyA signal: AATAAA"
complement(54977)
/note="TATA: TATAAAG"
55518
/note="polyA signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb
and 0.6kb early RNAs"
complement(55982..56935)
/note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27
in VZV and HFRF2 in CMV"
complement(55990)
/note="polyA signal: AATAAA, 3' end of 2.3kb and 1.1kb
early RNAs from 58568 and 57081"
complement(56132)
/note="TATA: TATAAAG"
complement(56948..58525)
/note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in
VZV and HFRF1 in CMV"
/codon_start=1
/db_xref="PID:g1334849"
/db_xref="SWISS-PROT:P03184"
/translation="MAHKVTSANEPNPLTKRLSSCPTRSGVTEVAQIAGTRPKMED
FVPTVONLKSQFAVGLLMAHSYLPANAERGIAYPPLVHTYESLSPASTCRVCDLLD
TLVNSDAPVAFEEDYALCYCLNAPRAWISSLTIGTMDFLHILIKYEPMAGGLDSLF
MPSRLAIDIQLEHYICRCFLPYSSSDMIRNANLYKLEFLKSLTGQSPANCFKS
MWPRTTFFLLPGPRTCKDSQDPVGDVIRNLTALCCLPTRNNVQHPFLRAEKGGL
SPEITRADYCGLLGTWQDGLLGGPGHHAIGLNAEYSGDELALALAITRPEAGDH
SQGPCLLAPMEGLRHKNARTICPLCESLGAHPDAKDTLDRFKSLILDSFGNNKILD
RIVFLIKTNTLDVPCPRLRAWLQWCTPQDFHKHLCDCPLCAINHSTNPVSLFGOI
YPPSFOAFKAALAAQONLEQGVCDLSLTLVLVIFKSTQVARVGKTLVDVTLELDVLR
IHGLDLVQSYOTSQYV"
complement(57081)
/note="TATA: TATTAA before BFLF2; BFL2 promoter gives
1.1kb early RNA"
complement(58088)
/note="TATA: GATAAAA"
complement(58568)
/note="TATA: TATTAA before BFLF1, BFL1 promoter gives
2.3kb early RNA"
58832
/note="TATA: TATAAAA before BFRF1"
58891..59901
/note="BFRF1 early reading frame, 1 NXT/S, homologous to
HFLF4 in CMV"
/codon_start=1
/db_xref="PID:g1334850"
/db_xref="SWISS-PROT:P03185"
/translation="MASPEERLLDELNNVIVSELCDSLSLEVERCSGAHVFSRGSOP
LCTVKLRHQIYHLEFYKFLAFKKNKNYESSPVFVINSNGLATLTCFHFPSGLR
SQSGPCGLSTLDVLPKNSIMLGQDDDFIKFSPLPVPEALDLKSVGVCRVITEH
RTWQFLVFOAANAQKASRVMDMSQSLRSQGVQEDTGARTVGGGPGPVGTHSG
CLGSHVGRGSGWLDNFSAEATEDEASYPWRDKDSWSESEAPWKELVRHPITRRH
RTETRRMRGSHSRVHPETRETIVGVGANRYSWRATPYLARVLAVTAVALLMFLR
WT"
59808..61583
/note="BFRF2 early reading frame, homologous to HFLF5 in
CMV"
/codon_start=1
/db_xref="PID:e25058"
/db_xref="PID:g1632788"
/db_xref="SWISS-PROT:P14347"
/translation="MALFLARHTLSGTGAGCHGRGPADPVSVDLTTLQALGERGFSRL
LDLGLACLDLSYVEMRFVWGRPPASEAAVTPGSLFRSHSAYLWLSVEYRPGGLV
RWARSQTSPLSLTAPLHGLPSLLSVVTGGCGGCAVAFCAFFLAYFLVVRSPAFAS

```

DRTAAMICDRSPCEINTRAVALRGYGLVKRFLAVFERSYDPPLLRQNRVPERCFA  
 IKNYVGLDSQSCVTPFSRWAQSHASELDREIRVTPTATPFAHADHASALIAS  
 LOKKADTCPCNPIOMWRTLVNSCLNSAHLPLIPAYSEGGKGGVGBELVAG  
 GYCLNSRDVFAIVSRNVLSCLLNPAAGPRAYKCFRSHASRVSGDPYPLAFCMDQ  
 GYCLNSRDVFAIVSRNVLSCLLNPAAGPRAYKCFRSHASRVSGDPYPLAFCMDQ  
 LAEPSSARSGWRPRAVLPNNALYELDRGSRLLDAILICLGDPRTCMRPVVLGVTV  
 RQTYLTLREARVAGSICQORQAPEDARDEPHLFSCLLEVELPGGRCAGCLRYQTR  
 YGTPAAAHPPGGAGGFCRQSPAS"  
 complement (61062)  
 /note="TATA: GATAAAA"  
 61344  
 /note="TATA: TATTATA before BRF3"  
 61507..62037  
 /note="BRF3 early reading frame"  
 /codon\_start=1  
 /db\_xref="PID:g1632789"  
 /db\_xref="SWISS-PROT:P14348"  
 /translation="MARRLPKPTLQGRLEADFPDPLLPKFQELNQNNDPNDVFRFAQ  
 RSYLVLTQFCVYEEVQRTFGVPRQRADIKRQASVAGAGAHHLGGSSATPVQQA  
 OAAASAGTGAALASSAPSTAVAQSATPSVSSISSLRATSGATAAASAAAAVDTGSGG  
 GGOPHTAPRGARKKO"  
 complement (62068)  
 /note="polyA signal: AATAAA"  
 62069  
 /note="polyA signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4,  
 3.1, 2.5 and 0.8kb early RNAs"  
 complement (62078..71527)  
 /note="BPLF1 reading frame, 1 NXT/S, analogous to VZV  
 RF22"  
 /codon\_start=1  
 /db\_xref="PID:g1334853"  
 /db\_xref="SWISS-PROT:P03186"  
 /translation="MSGMDGQSQRGTGVPVGIITMDVNPAGCGGSGSALRIILGTA  
 SCNQACKRGFRAGICVSNCLYLIVKSLFAGRLPLRSRDELDEVLEGARLDALMRQS  
 GILKGHEMAQLTDPSSVLRGGRGVHIYKSAEIFGLVLPFAQIANSAVVQSLAEVLH  
 GSYNGAQTILYICDJIYAGAIITETDGFYLEDPHCKDAAGTAAHVSQYSAEHL  
 GYVAGPAQVTCVHLFPEAFETEDPRIFLMFHYGVTFYFANGSGFDLVGLPELVSS  
 DGAAGTCPADSPPPVMLPFERRIIPYNLRLPSRSTSDSPFAARYSPAKTNSPPSS  
 PASAAPAASAPASAPASAPASAPASAPASAPASAPASAPASSPLFIPILGHTPT  
 GVPAPSTPRASSGAAPOTPKRKKGLGKDSPHKPTSGRRLLPLSSITTEDQLPRTH  
 VPPRPPSALPPVPIPIPHQSPASPAPAPATPIAPSVTSPTRPLQIPILPIPUQ  
 AAPSNPKILPTTSPATAAATPTTLTSPPTQQPQPSAPAPAPSLPPLQOQPTPSA  
 PAPSLPQQQPPSAARAPSLPQQQPLPSATPAPPAQPLPSATLLEPKNHP  
 PADRAGTETSPSPGQOQPSFGDASGGGLVYLSLDEEPFLMSDEEAESDLAS  
 DIPPTDEDMFDEVEFSNLESQSAPTSPITLDTARSOYQOTFIDIEPMDFVPLE  
 SNIRTAGHTYQOALVYDPASNRVEPADALSMDIYLLVTVVLEOGLIRSRDRSVL  
 NLDEFUKSGHVLPDLDQLTSELNQNANLMSKNGRGEFHHKLAALKLEAC  
 LPSLATKAVRVDAKMLAEIPQLAESDDGKFDLEAARRRLLDLSGGDQAGEGGG  
 EPEDNSYGRGVDPVPLVLDDESCKRLLSLAEARTAVARQAQGVDEEDVRFLLALTA  
 IEYGAPPAASVPFVHNVAVRSKNAALHVRCTADIRDKVASAASDYLSEDPISLPT  
 VMDFDLTLHRTCOILASLPLNIRYTSIEWDYRELLYLGTLALSDMSGIPWPLERY  
 EMDPSIAPLPERETVAKKQKELETTRENEKRLTILDIEMALGLAGVASAPGAPIS  
 EMDPSIAPLPERETVAKKQKELETTRENEKRLTILDIEMALGLAGVASAPGAPIS  
 PAPSPTANHNDPEATVPPLADTAALTIPVIEKITYANAGSIVGAANKNPIYIRLDTIQ  
 QVYRSKYLMLNKLSTFTFYIDNYIASFEESIDHLYRDLPLVDPEQDGDITGMPV  
 SEALHFFEMGNRLTLPARVALQNEFATHSTLKETAAYANLPLGLLADATITGAP  
 EDALRLSGLONLSOTLIPGKLRFLSVLOKKNNDOLROKQOAWRLAEQEFK  
 PATEOLOELFVAPNKLKROYEKKIROLMETGRKEKLEQEDKEQERREARAN  
 EAWARKALGARPEAPSPDWNLTLLASLLPDNTDSAAAAAANRNTDILDSLTQ  
 ILAALLGTLVRRERRLSLLVDGGAERMAEAPGFTDITETGLARLDAMPAPPA  
 ATAKEGGGRGAEAGALFRATAADAIRSALAQTROALQSPDKMSAYVNTDLEAPY  
 AVEYERGLAGLLEKRAEAALTAIVSEYVDRTLPPEATNDPQANLPPTPIQATAPM  
 RLASDLSALPKPKPOLLTRRDLQATGDTFFSELLTEAAEAEREALEREQVESOTLM  
 AKAHENASTRRGFHTALEAVLSRDEAPDELRSLLSPPKAPVQAPLEAARAA  
 AGNSWPKYKSLAAAKWICICAEVRGLSEGALAGAGAGANLNLAAADGETHELTR  
 LLEVEGAQNSMEELRLALATLDPKRVAGGETVADWKRLSLEALILIOEAQES  
 QOGLTQDLVQARTDTPRQLKIVVEAARGALGASAGSAYALLKDKLLIAAQAQES  
 FLAFYTAQTVFKHPLTNPLLTISAPPTGNGAPTRAQFLAAGAAKAYAGT  
 LWLETSPCDPLNPAVVSADTQELNP IPVYHNFLVMTYVLENPEAFSLPAGRPQ  
 LWLETSPCDPLNPAVVSADTQELNP IPVYHNFLVMTYVLENPEAFSLPAGRPQ  
 AIGCPDODDERRTLASVASARLSAAADSWYDTPWPDVSNAGELLREYVSAPKALM  
 EDLADNPVYNTLLHASLIASRNHPYPAPATDREVILLEQREMMALLVGHYPAYAA  
 AFLGAPSFYAGLGLSALARDGGGLGSDSLVLYRLVRSFASGRGMPSTTSGNSDG  
 EDARRLTRHRIAGPPTGTFIFFOAWEEMDTRAALWPHPEFLGLVHNQSTARACMLL

LARRCFAPALQOQLWHSRLRLEGPVAFQDYLRDFYKQAYTRGEELEPRAEBLEVPRETP  
 SSGYVTVGRALNLMPYGTPTGPRKSGSDTIPVSEFAQAAVAAALGRRLTLFVFSQY  
 LFNLTQGRVVVALLYCDGHSPEPRLVETISLNFQDLQDGYSESPPEMSIFARQ  
 AYLRELLTEAPAAKPKARPPTVAILANKNIIMKCTYRINLNDVOYFYNAGASR  
 WPTDLNVSFYEHEDPPFVGQLPNERNVQELSGFPFPRHGVGLVSDGGFQSDANT  
 PASSROLDQGGGETDQGEKSTTAESAGPPSPQSPILKEKVGPRGMDLSPTSSP  
 RDVYTPGLAAPTILGPRLMARPYGTASTRASESPDRSPSGSPWPWDXSLLELLPQ  
 AQOQPPSPWASEQOPIVYTLSPHSTPSTASGOKHTIQIPGLVPKSPKSPPSAPY  
 PKQSTGTGATPTPASAALTTFGLQPDQTOASSQDPPYGHSMOREKKQOGGREAAEI  
 RPSATPLTAVGLRPRVAVAGAASATPAFDGCEAPSGFPIQAPALSGLAAPAH  
 TVGALARPQKTAQRQDAAALFTTIKAVGARPVPAKATGALAGAARPGQPTAAP  
 PSAASPPRVSLRQSQSPAIPLPMHSGSEGARPEVRLSQYRHAGPQYTVNRKE  
 APPSAASQLPKMKCKDSMYTPPSGARYPAFOALSFSQSVASHPADQYTLWNT  
 PSVWTFQFSIEDIIEVVTGSGTSGDLVVPSPSSSLSTAAPQODRLYSLTLQASRV  
 LSRFVSQLRRKLERSTHRLIADLERLKFYL"  
 62249  
 /note="BAM: BamH1 F/Q"  
 62430..62477  
 /note="Site III for EBNA-1 binding (Rawlins et al, 1985)"  
 66121  
 /note="BAM: BamH1 Q/U"  
 67477..67649  
 /note="Exon in EBNA-1 RNA (Speck and Strominger, 1985) and  
 cDNA clone T4 (Bodesscot et al, 1986)"  
 69410  
 /note="BAM: BamH1 U/P"  
 69684..69930  
 /note="5 x 51bp repeats"  
 70387..70521  
 /note="9 x 15bp repeat"  
 70750  
 /note="TATA: CATAAAA"  
 complement (71520..75239)  
 /note="BOLF1 reading frame, 1 NXT/S analogous to VZV RF  
 21"  
 /codon\_start=1  
 /db\_xref="PID:g1334855"  
 /db\_xref="SWISS-PROT:P03189"  
 /translation="NASAMESSGSGGGAQADPPLAEVDGGLARVTROLLSSGDDPA  
 ARLRALMLELIFGLDLOAPVLYRDELNTLTMGSHAYPAVALRHHAYLYLRASF  
 SRVFGHGLEAALVLAASSLPTTASPDATDLOGSRLIAETRALAAAYRIIEBGS  
 GVLVSGGTAAFAVEELVADTYLARMDAFREGLSFYAFNAKTILGRWLTVTAE  
 TNRYNAAAGQOQTAADIKAMAVELHSGGGAGGEGEGGGLFHRPESLSVVA  
 SLPLARRAAGVILGYVASGQTPVAAVPVLAFAARLRLLESPGALFYDYVYVLA  
 LDWQYGVDSVIEAFLAGMAEMEAALVOEAGSASFPAAEQVATVLLSAGL  
 NETVADGYAMLLASVPYRSRWRMLEATAALLESISGFALHFFLLPTASRPAR  
 VARAYLRAEAEDRRARTSGSTPAAAPATAATVAVGVAADPDWATVPLRIFVPP  
 PAAYEQVAGDUSSELLSLWVYRSLWQAAPAPALPCKPPLLPGEGRROWTAAY  
 AAAPTIDVAYCRSLRAGTARADPAYVHSPEFAATFEQIWPALRRYLSNELPKTR  
 SLAALRWLYFGSDIALSPELTRARPLIELIYATVWEIYDGAAPPMPGSPQAVGLRP  
 LNLGEGKAGDAGAEAGEDEGGPGWGLSHDAVLRIMDAVREVSGIISETISASERA  
 AEAPLWPTSLFLLTLRYSTTAESLGLATRELVSGETLSIEDISRLTGAARLCS  
 RPLLYDAGTGRVQIPLAEEEEAVVAKESVSPRHSYSTDLOTLKASVVEGIDQVC  
 ROAAALATADATLRRRLPALRESGIDHPLWAHTSEPURDUEELNERVEH  
 ALEJESLGTALRSVAYFRDYTFARLQPPAIDAERAELVRRDARDPPVFPAPR  
 RLQOGAGTTPPLSMDLILYKSKICALVDLHDHPAPETPTIKTYPAMDNLPEQ  
 ITVTPRSVLAFAARTARVOTHLVPALTDDSPSPVGTOPPPFRILPAKLAAILLG  
 NGNASKRARSDDLSPPHGRWALVDSFSSDSFSSDSDODEGEGEADLGRVGGG  
 GGAEEDRSPDIDTARAOKVETSCPRRSPTTPSPSRASGGGGPDRGEAEAH  
 TYPPLSAAAAASRVPRTRRGATRRPRPTAEDE"  
 complement (72192)  
 /note="TATA: TATTAAA before BPLF1"  
 73468  
 /note="BAM: BamH1 P/O"  
 75017  
 /note="TATA: TATTAAA BO-RI late promoter before BORF1,  
 gives 3.9kb late RNA"  
 75238..76332  
 /note="BORF1 late reading frame, 2 NXT/S homologous to VZV  
 RF20"  
 /codon\_start=1  
 /db\_xref="PID:g1334854"  
 /db\_xref="SWISS-PROT:P03187"

```

/translation="MKVQGSVDRRRLQRIAGLLPAPRLNISRGSSEFTRDVRGLVE
EHAQSSLSAAAVWRAGLAPGEVAVAGGSGGGSFSGSWRPPVFGDFLIHASSFNN
AATGPTLQFQKQSPFSGVDVFTPLSLFILNHNHGRGVAARVEAGGGLTRNALLIYD
SPATLADLVDFGRVADRRFNFITPVGPLVENIKSTYLNKRTITTVHGPVYSKAIPR
STVKVTPQEAFLDADLWLSGGAGGGGVCVFGGLQPCPADARLYVALYIEEAGPR
FTFFQSGRCHCOIMLRIYYSPSIHRVAVVQPLHIELTGTGACVACLGTFSATDGR
RFAFNVRGSLPVEIDSPYSNVSDWEVL"
complement(75322)
/note="TATA: TATTATG before BOLF1"
75819
/note="TATA: TATAAG"
75838
/note="polyA signal: AATAAA"
complement(76126)
/note="polyA signal: AATAAA"
76169
/note="TATA: TACATAT BO-R2 early promoter before BORF2,
gives 2.8kb RNA"
complement(76300)
/note="polyA signal: AATAAA"
76407. .78887
/note="BORF2 early reading frame, 2 NXT/S. Homology HSV
140K ribonucleotide reductase (Gibson et al, 1984) and RF
19 vzv"
/codon_start=1
/db_xref="PID:g1334856"
/db_xref="SWISS-PROT:P03190"
/translation="MATTSHVEHELLSLIDELKVKANSDDPVLAGRLLHRLKAES
VHTVAEYLFVDFYDEDFEOMHRLDETRVSFAFQSPAYERIVSGYLSALRYD
TLYVGRSQKQSOVHFYMRLAGFCASITCLVAGLRAALORARPIESDMVEFDYFE
HLTSYVCCSTPFMPFAGVENSTLASCILITFDLSEWDVQALRTHGSLDFQRAG
GVGTAGQDGRHISLLMRMINSHVEYHNIGCKPVSVAAVMEPHSOIFKPLETKLP
ENHERCPGFTGLFPELFLKFRDTPMSDWLFDPKDAGDLERLYGEFEFEYRLV
TAGKFCGRYSIKSMFISVNCVAKGASPFILKKEACNAHFRWDLOGEAMNANCAEV
LQPSKRVATCNLCIPCLVNAPLAVRAORDTOGDELLALPLRSVLTPGGVAY
GDGFSIARLATOCACTFVACSILOGSPTVDSRDMASNGLVQGLADVADLGHQYT
DPPSRLNKEIFEHYFTALCISLIGLHTRKIFPGKSGTAGGFHWHWDAGDLS
LPRELWSRISERVIRDGFNFSOIALMPTSCGAQVTGSDAFYPYANASTKVTNKEE
ALRPNRSWRHRLDREALNLVGRVCLPRLQRYLRFTAFYDQEDLIQMSRD
RAPFVDQSGHSLFREDEAARASTLANLLVRSYELGKTIWYCRIEKAADLGYMEC
KASAAALSVPREONERSPAEQMPRPMEPAQVAGPVDIMSKGPGSGPGWCVPGLVEY
CYKXROLFSEDDLLTDFGTERACESQ"
77835
/note="BAM: Bam H1 O/a"
78804
/note="TATA: TATAAGT Ba-R1 early promoter before BarF1,
gives 3.5kb RNA"
78883
/note="polyA signal: AATAAA, end of 3.9kb late RNA from
75017 and 2.8kb early RNA from 76169"
complement(78896)
/note="polyA signal: AATAAA"
78900. .79808
/note="BarF1 early reading frame, homologous to HSV 38K
ribonucleotide reductase (Gibson et al, 1984) and RF 18
Vzv"
/codon_start=1
/db_xref="PID:g1334857"
/db_xref="SWISS-PROT:P03175"
/translation="MSKLLYVRDHGFGFACLVETHRNWFAAHIVLTDKDCGLKLLNE
RDLFEYKLTFLAAKELVNFDIELVTSFESHDIHYTEQKAMENVHGETYANIL
NMFPGDRAAMNAYAEAIMADEALQAKISWLDRKAAAVTLPKTLVFLLEIGIFFIS
SNFYALLRVRLGMPGICLANNYISRDLELHTRAASLLYNSMTAKADRPRTATWQELF
RTAVGEVATFIEARGEVTLVDVRAIKFLEATADRLIGDQIQAPLYGTPPKDCPLT
YMTSLKQNFEEQSSDYTMLVVDL"
complement(79495)
79537
/note="TATA: TATAACA"
79537
/note="BAM: Bam H1 a/a"
79840
/note="TATA: CATAAAT BM-R1 early promoter before BMRf1,
gives 2.5kb RNA"
79899. .81113

```

```

/note="BMRf1 early reading frame. Early antigen protein
recognised by R3 monoclonal (Pearson et al 1983; Cho et
al, 1985a)"
/codon_start=1
/db_xref="PID:g1334858"
/db_xref="SWISS-PROT:P03191"
/translation="METITQLRFKTKALAVLSKCYDHAQTHLKGVLQNLVLSVNYGG
PRLAAVANAGTAGLISFEVSPDAVEMQNHQSPEEAPAAVFRNLAYGRTCVIGKELF
GSVEQASLOFYKRPOGGSRPEFVKLTMEYDKDKVSKSHHTCALMPFMPASPDLRNEQ
MIGOVLLMPTKASSLOKWARQGGVKVTLNPDLYVTYTTSGEACLTDYKPLSVGP
YEATFGPVAKQDQAGVAEHVVCVAAADSLAALSCLRIPAVSVPILEFYSGGITAVP
AGLTSAGDPLDLISLVILENHAEEAASASTASEPDKSPRVQPLGTGLGQPRRHVSP
SPSPPTPTPTWESPAPETPPSPAIPSSHSNTALERPLAVLARKRTTSSEARQKQKH
PKKYQAFNPLI"
80779
/note="TATA: TATTATA BM-R2 late promoter before BMRf2"
complement(80782)
/note="polyA signal: AATAAA"
80832
/note="TATA: CATAAAA, possible promoter for 1.4kb late RNA
encoding BMRf2"
81118. .82191
/note="BMRf2 early reading frame"
/codon_start=1
/db_xref="PID:g1334859"
/db_xref="SWISS-PROT:P03192"
/translation="MFCKOHLSCAGVCFLGLLASTPFIWCFVFNALLSLEIFSPQW
THVRLNGPTACLMAVLTLPYAKHAVRAVTPALMNLIASLIFSLRVYSTSWSA
CFLFANLPLCLPRLAIEIVICPAHQRFELGLLACTIFALSVSRALEVSAV
FMSPFIFLALGSSLAGARNQITSLGRRSIFCARGDHSVASLKHETLKPCWDL
LAISALTLYVVCVMVLHVHAEVFFGLSRPLFLFCGAMSGGLYHSHSIIACVMAT
LCTISVYVFLHETGLVGLTFLVIFISIFYFVFGVAALSAMRYKLKFKVNGPLVHL
RWVMCCFVF"FCSEVLLVTFIKS"
81751
/note="TATA: CATAAAT"
82180
/note="polyA signal: AATAAA, end of 3.5kb early RNA from
78804, 2.5kb early RNA from 79840 and 1.4kb late RNA"
complement(82311)
/note="TATA: CATAAAT"
82319. .82461
/note="2x71bp repeats"
complement(82743. .84059)
/note="BMLF1 early reading frame. Diffuse early antigen
(Cho et al, 1985b). Also homologous to RF 4 vzv and I563
of HSV (BSLF2 + BMLF1) is also called EB2
(Chevallier-Greco et al, 1986). General transactivator of
transcription (Lieberman et al, 1986)."
/codon_start=1
/db_xref="PID:e25042"
/db_xref="PID:g1632790"
/db_xref="SWISS-PROT:Q04360"
/translation="MEGEESTDGEISSSEEDDPTPAHAIPARPSVVITPTSAS
FVIRPKWDLQDVTTLHRSPLCDEDEKEETGNSSTYRHKRRGVEHGTGDSYSG
RHLPFGARAPRAPRAPRAPRAPRAPRATGRPSRSGRAGSTRQAQERS
ORLPKAPFMDMSLVKPVSKITFVTLPSLASLTLEIOPDFLOSLMAVAHAPEIGAW
QKQVPHELRRSYKTLREFFKSTNKDTLDMQAIQONAGICTIVAMLEETIFWLOE
ITYHGDPLAPADEILLACAMSLSKVILTKELAPCPPLNTRDYNFVKQIFYTCAT
ARQKNVETLSSSVKQPLCGLAAVAAVPAIINANRRRHDEVEFLGHIYIKNNPGT
LSSLTAEVETHTRDCRSCSRVLRAILSPGTGSLGLFFVPGLNQ"
complement(82747)
/note="polyA signal: AATAAA"
83640. .83729
/note="10x9bp repeats"
complement(84122)
/note="ACCEPT: CTCCTCTGCAG acceptor in spliced form of
BMLF1 RNA"
complement(84227)
/note="DONOR: CAGGTAGA donor in spliced form of BMLF1
RNA"
complement(84229. .84288)
/note="BSLF2 early reading frame in 5' exon of spliced RNA
encoding BMLF1"

```



```

CDS
complement(89567..90013)
/note="BLF2 early reading frame (BLF3 in Baer et al, 1984)"
/codon_start=1
/db_xref="PID:g1334867"
/db_xref="SWISS-PROT:P03199"
/translacion="MCPVROHPAQAPPAKFAQLETVPHPONRGMLSPKAPPKMQR
RPRPVAKRRRPSQOVRPILPVVESTQDMFGVQSPFOITAVIOLQRDRDM
RPPIYFALLANCAGPAGLRAHRLFPQKPPCCQSRFSPDSQTSFC"
complement(90051)
/note="TATA: TATACAA BL-L2 early promoter before BLF2, gives 0.7kb early RNA"
complement(90062..90652)
/note="intervening sequence in gp220 gene"
90177..90639
/note="21 copies of 21bp approximate repeat"
complement(92192)
/note="TATA: TATTAATA BL-L1 late promoter before BLF1a,b. Gives 2.8 and 2.2kb late RNAs"
92238..92581
/note="Exon in Bodescot et al (1986) RNA (spliced from 20763 to 92670)"
92243..92602
/note="BLF3 reading frame"
/codon_start=1
/db_xref="PID:g1632791"
/db_xref="SWISS-PROT:P03202"
/translacion="MDKRGPPALDDNMBEVPSTSVVQVQVSGAWENVLIELSDS
SSEAEADAHLEPAKQKTKRKRRVDHDAGGAPARPMPLPPQDLPGRALIRRFPLDLR
TLLOAIGAAATVSIPIA"
92670..95248
/note="Exon in (Bodescot et al, 1986) RNA from 92591, to 3' end"
92670..95162
/note="BER1 frame, homology with BERF2b and BERF4. A fusion of BLR3 with BER1 encodes ERNA-3a, latent cycle gene. (Hennessey et al, 1986, Joab et al, 1987); Protein sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25053"
/db_xref="PID:g1334871"
/db_xref="SWISS-PROT:P12977"
/translacion="RIDTAIDQFGSOISNTEMYIMYAIRQAIRDRRRNPASRRD
QAKRLQTLAAGWPMGQAYSSWMSYIDHQTTFVHLQATLGCTGGRCHVTFSG
TKLPRTGPDQRLYQVSSGNIVQCNPRYSIFFDYMAIHRSLTAKIEWELTPDQR
VSFMEFLGRLTDLSTYKFSVSDALGTSIQTPWIDNPTETAQAWNAGFLRGRAY
GIDLLREGEVEGATGRTSESDTESDGDDELPCIVSRGPKVRPPTFIIRLLHR
LLMRAGKRTQGEKLEKARGSTYGTPRPVKPRPEVQSDSTATSHGSAQVPEPP
TIHLAQGMAYPLEHQGMARCPVAQAPPTPLPVPSGDLPGVFSGRVACAPVAP
AGPIVRPEPSTLQAQAFAPVRPQHMPEVFPVTVVALERVPYKPRVPAPKIAM
QCGPGETGIRRARERWRPAPWPNRPSQMSVRDLRLARAEQVQASVEVQPPQ
LTVSPQPMEGPLVPEQMPFGAPFQVADVVRAPGVPMQPOYFDPLPIQISOGA
PVAPLRASMGVPVPATPOYFDIPLTERINOGASAAHFLPQPMESGPLVPEQWMPF
GAALSQSRPVASQVFDLTLQPINHGAPAAHFLHPPMEGFWPEQWMPQCAPPS
QGTDVQHLQDALGTLHGLNHPGVPSVAVNQLHQAFLGFLDESEGGSDTS
PCEALDLSIRGCPQAPVQVPEEGQDQATEVLDLSLTHGRPRPTPEWVQVGGGON
VTGPETRRVYVAVVHMCQDDEFDLQDPDEA"
92703
/note="BAM: Bam H1 L/E"
complement(93161)
/note="TATA: CATAAAT"
93479
/note="TATA: TATAGA"
complement(93482)
/note="TATA: TATAAAT"
94208..94277
/note="repeat type A"
94281..94306
/note="repeat type B"
94307..94381
/note="repeat type C"
94386..94411
/note="repeat type B"

repeat_region
94412..94489
/note="repeat type C"
94490..94560
/note="repeat type A"
94571..94648
/note="repeat type C"
94649..94719
/note="repeat type A"
94896..94982
/note="repeat type D"
94983..95069
/note="repeat type D"
95221
/note="polyA signal: AATAAA"
complement(95272)
/note="polyA signal: AATAAA"
join(95353..95709,95788..98247)
/codon_start=1
/product="EBNA3B (EBNA4A) latent protein"
/db_xref="PID:e276431"
/db_xref="SWISS-PROT:P03203"
/translacion="MKKAWLSRAQQAADAGGASGSDPDYDQGNVTQVGSPEISPEI
GPFLSAASEDDPQSGPVENLDAAAREEPEHEQNGDDPLDVTQRFRFVDVNP
TQAPVIQIHAVYDSMLQSLPLGSLFLLEONLNIIEFIMCMTRHRCQAIRKKPLP
IVKORRWKLLSSCRSMGYRTHNLKYNVSFSGDGNVHPLVLTATLGCDEGTHATY
SAGVQIPRISDQNKIETAFLMARRASLSAERYTLFFDLVSSGNTLYAIWIGLTK
NRYSFIEFVGLCKKDHTRWFROCTGRPKAAKPLWRAHPVAIPYDDPLTNIELIDL
AYARGAAMNTEAPRLPDDPLIIVEDDDSEIEAESEDEEDKSGMESLKNIPQTLPYNP
TVGVRPAVDRKSDAKSTKKKRALIVDFSVIKALEEHRKKKARTQPRATPSQAP
TVLQRPPTQEPGPGVPLSQAARLEWQPLPGPQVAVLLHEESMGVGVHSGMLDL
LEKDMEQREVRMATLLPVPPQPRAGRGPCVETGDLGIESDEPATEPVHDQLLPA
PGDPLLEIQLTSPITSQLSSAPSCAQTPWVPVQSQTPDDPTKSRPPTAAAPROW
PMPLRPIPMRPLRMPPIFNHVPVPTPHQTPQVETITPKYKTAOIGHIPTYOPTGPA
TMLLRQAPATMOTPPRATPMSPEVPVPRORPGAPPTPPQVPPVPRORPRA
PTTPPQVPLTPMQLRALRAPAQOQGTQKILROLLGCGVKKGRPSLKLQALBRQA
AGWQPSGSGTSKIVQAPIFYPPVLPQIQVMGGGSPATAAASAVTOAPEYTRER
GVGMPGLTDIIPPSKRAKIEATPEPMHGGASHSPVVLNVNGGQQQTLECGGTAKQ
ERDMPLGLDIIVSSPSSSETSNDE"
95353..95724
/note="BERF2a reading frame"
95725..98244
/note="BERF2b frame, homology with BERF1 and BERF4. BERF2a and BERF2b are spliced together to make EBNA3B (EBNA4A) latent protein."
complement(95819)
/note="polyA signal: AATAAA"
complement(95853)
/note="TATA: TATAAAT"
complement(96276)
/note="polyA signal: AATAAA"
97522..97698
/note="3x60bp repeat"
98233..98769
/note="BERF3 reading frame"
98364..98730
/note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"
join(98371..98730,98805..101423)
/codon_start=1
/product="EBNA3C (EBNA 4B) latent protein"
/db_xref="PID:e276455"
/db_xref="PID:g1632793"
/db_xref="SWISS-PROT:P03204"
/translacion="MESFEGGDSRQSDNERNVDVOTGTFHDODPGGPPSSGASER
LVEESYSRDQOPWQSGRQDENRCWMQIRRRRRRAALSHLLDLDENVPDWRPPH
ITPYTARNIRDAACRAVKQSHLQSLNILDGLDQTHILCFVMAAQRQDLIRBPL
VAEGVGWRHMLTSPSQSWPMGYRTATLTLTPVNRVAGDSIMLTATFCQNAART
LNTFSATVWTPPHAGPREQRYAREAEVRLGRKQRYRYIDILICLSLHHIWN
LLOTEENLLDFVPMGVMSCCNPNVNFHKTIGNFKPYVPWNPENYPHARRGI
KEVIONAKRQIQISMLATGEPGSDATSETSDDEDTGROGSDSDELDELPI
IDPMEVQQRQPVNFVSRVPAKPKPLPWTPTKTHPVKRNKTSRSDSKAAQSTPE
RPGSESSVTVPEPAHTVILHQPVPVPPKPTPPPSRRRGACVYDD
DVIEVIDVETEDSSVYQPNKPKRHQDQGFQSRGRKRAAAPTVPFSDTGPVAVGP

```







```

/misc_feature
112620 /note="TATA: TATATAA"
112620 /note="BAM: Bam H1 K/B"
113876 /note="TATA: TATTTAT before BBRf1"
113876 complement(113885)
114204 /note="TATA: CATAAAT"
114204 /note="BRRf1 late reading frame, homologous to RF 54 VZV"
114204 /codon_start=1
114204 /db_xref="PID:g1334884"
114204 /db_xref="SWISS-PROT:P03213"
114204 /translation="MFNMNVDESAGLASSAIPVHPTPASVRLFEILQGYAVVQGG
TIYANLRNPGVTSROVFTHLKRAISHCTYDDVLHDWNKFEACIOKRWPSDDSCASRF
RESTESWSTTKLITNTIYVLRHSVLSYERVDVICATGMVPAVKKPTIO
ELHSAIKSLDRVCVRELGHSTIRISIGTELZEAKELIESNSFIQFTVETIYL
LNPATVATIGGRRRLHVFPFAIFAGVTFDSVPQDIQNIWFIRLEHAKICQL
KSTDELAIVGGGRDMYKDILAHLEQNSQRDKPELLNLLKISENKTISGTDV
EEFITDASNNLVDNRNLFQPGETAAGLKKRVNTVVKCLDQINEQDFQINGEKE
RELYLKRSMESQLQASLPGGNPAASAPAAVAASVDILGTGASATKLFNS
PSASLGRVSGHNESTLNSFSQYIPPSREMTKDLTELWESLENTFKLTPVVDNQGQ
RLYVRSYDPTISILGPFYVLAELSPVELYDVATIGIVEILIDELVSSRLAIYE
DLGRKCPASATGGGHHGIRQAQFSAGDTEPDHAKSNPARDPPPGAGS"
115948 /note="BRRf2 late reading frame, homologous to RF 53 VZV"
115948 /codon_start=1
115948 /db_xref="PID:e25024"
115948 /db_xref="PID:g1632797"
115948 /db_xref="SWISS-PROT:P29882"
115948 /translation="WASGKHQPGGTRSLTMQKVSRLVTPRLVLEVRNHNATCVATNV
PFYNARGDNLTRDAHVYKARMSISQFCGYVLVSLDSEDQVDHNLIPPHVFSERMI
LKPNNVNLMETAMKSIENAKSIPGICREVGLRLTLHSCNNLDSLFYNGART
LLSTLVKYHDLREGAATPGPWNEGLSLPKLHKLKRAPSEARDLQSLFLTSGKMGCL
ARSPDYCADLKEEDANSFTENLFYQDSLLTKHFQCTVLTQLRRKCLGSDTVSKI
IP"
115948 complement(116683)
115948 /note="TATA: GATAAAA"
115948 complement(116696)
115948 /note="polyA signal: AATAAA"
115948 complement(116784..117386)
115948 /note="BBLF3 early reading frame, spliced to BBLF2. BBLF3
contains a consensus nucleotide binding site; Protein
sequence is in conflict with the conceptual translation"
115948 /codon_start=1
115948 /db_xref="PID:e25025"
115948 /db_xref="PID:g1334887"
115948 /db_xref="STREMBL:Q04343"
115948 /translation="AFLOQVKDSEASRLDRDVMGGEATVARRHIVKARRPGGCLLM
AIFQGLYVGGREHSGPFLVWEAFSWTLQDLAARPEADKAPPDHDHLLTLVRDLTR
RLAPGRRNRWALPRAWLQRLRAGLRSLGSHVCLLDKDGARPCQTATEHGLSPT
AYFREIMAFLLDVISALHPGYTIPMETITRETDLLMTVLSLF"
116785 /note="polyA signal: AATAAA"
116785 complement(117386..117515)
116785 /note="intron spliced out in RNA linking BBLF2 and BBLF3"
116785 complement(117515..119080)
116785 /note="BBLF2 early reading frame, spliced to BBLF3;
protein sequence is in conflict with the conceptual
translation"
116785 /codon_start=1
116785 /db_xref="PID:e25026"
116785 /db_xref="PID:g1334888"
116785 /db_xref="SWISS-PROT:P30118"
116785 /translation="PDVLKGPVLLRSQTMETPAESYARVSVTYFYNVTQAGRWAA
INWGIPIKREDVEFLIVVQCPPLGSGLEPPVNAFSTTELFLNWERELRRSGG
LIAMADAAEKDLDSFRTRDRLLLSAARVEDEGLIFQPLFPAAQVVCQSGDGG
DQPPFDVFGEGMEQCTCPAQRHSESPGQLDVYIRTPGDVFTYETETDDPSVP
PFRLDILPYTEVDLVSDGATRGGDARRHVSRLKILEPAGDFESLWNSWMAAGG
LVAFRLSVASCVANHRGKPIFYLLDPELCPGSGDFQYVPGFPLPHYVGRAPA
FVHRAPHSGLLLDLNLGVSGTPLADALLGLDARSQGRGSLILQIQIWPPTKKEINP
RVIVREGSGEGDETTIVVGRGAETIALEADATNWLVIELARCHISARGAPVGTDPGG
GQARDAQWLRLHVGTSRTRALGGGLYATVTRVLLHAAADLGLTWAYADEFLILFGV
APTSAPHSPEPLAQ"

```

```

promoter
118981 /note="TATA: TATAAAA BBR1 late promoter before BBRf3"
119067 /note="TATA: TTATAAA BBR2 late promoter ?"
119098 /note="TATA: TATTTAA BBR3 late promoter before BBRf3"
119108 /note="DONOR: AAGGTGAAT possible donor"
119137..120354 /note="BRRf3 late reading frame"
119137..120354 /codon_start=1
119137..120354 /db_xref="PID:g1334889"
119137..120354 /db_xref="SWISS-PROT:P03215"
119137..120354 /translation="MKSSKNDTFVYRTVKTLVTVYVFMVMSAVVPITAMFPNLGVPC
YENALVDYGAENLTNYNLAHLLPTLYLEPEPEMYITLYFIADCAVFIYVACGEVAL
IKARKVSGLDLSNAWVAVGSPVTFLLAILKLSIQFIQVLSKHVLSFAVYFLH
FLASVHACAVTREFSPVWVKAQNSIQDFTLWVYFLKPVVTVNLVGLCLAEFL
VFSLSVIALGNSFYMGVMDVLANFLILPIFIWYILTEVMAILCFIRFKMGFCGM
FIASITILPLVRYEAVFVSAKLHTTVAINVAIIPILGSVAMLIICRIFKMGRCOTD
YVPSVETVELESEPRSPRSPNTSPGRNRRSSTSSSSSRSTRQRPVSTQALVSSV
LPMVTDSEEIFP"
120260 /note="ACCEPT: ATCTTCTCCAGGT possible acceptor"
120260 /codon_start=1
120358 /note="polyA signal: AATAAA"
120358 complement(120747..120974)
120358 /note="BBLF1 late reading frame, possibly homologous to RF
49 VZV"
120358 /codon_start=1
120358 /db_xref="PID:g1334890"
120358 /db_xref="SWISS-PROT:P03216"
120358 /translation="MGALWSLCRRVNSIGVDGCIINLYNDYEEFNLETTKLIAAE
GRACGETNEGLEYDESDENDELLFLPNKKPN"
120358 complement(120764)
120358 /note="polyA signal: AATAAA, 3' end of 0.6kb late, 1.6kb
early, 3.0kb early RNAs"
120358 complement(120929..122341)
120358 /note="BGLF5 early reading frame, homologous to RF 48 VZV
and alkaline exonuclease of HSV"
120358 /codon_start=1
120358 /db_xref="PID:g1334891"
120358 /db_xref="SWISS-PROT:P03217"
120358 /translation="MADVELEDFDMEEMTSYTFARFLRSPETEAFVRNLDPRPQMPAM
RFVLYCLCKIOEFGFCDYVSSLVQENDSKDPSLSKSIYWGLOEATDQRTVL
CSYVESMTRGOSLENMWDILRNLIISSSKLLSTIKNGPKVPEPAPISNHYEGGVA
FGLRCEDTVKDIYVKLIGDASANRQFGFMISPTGICFVGLDLCVNVESQGDILFT
DRSCIEYKCRFLYFSSEFDPIYPTALYKRCRKSFIIRINSIARPTVEIYVPGD
RLPSEGDXLLTQDEAMNLDKVRKRLGPHDLVADSLAANRGVESMLYVMTDPSNAG
RIGIKDRVPVNIIFINRHNIFYQLQYKIVGVYVHSGGKPGRDCSPRVNITVAF
RRKSLPDATCTLGSDDLDDASVEIPAVLVTPVVPDVSIRKTLSTLAAGSNKAYADN
TFDTAPWVPSGLFADDESTP"
121669 complement(121331)
121669 /note="TATA: TATTTAA BBL1 late promoter before BBLF1"
121669 /note="TATA: CATAAAT"
121697 /note="TATA: TATAAAG"
121772 /note="TATA: CATAAAG"
122313 /note="BAM: Bam H1 B/G"
122313 complement(122328..123692)
122313 /note="BGLF4 early reading frame, homologous to RF 47 VZV;
protein sequence is in conflict with the conceptual
translation"
122313 /codon_start=1
122313 /db_xref="PID:e25030"
122313 /db_xref="PID:g1334892"
122313 /db_xref="SWISS-PROT:P13288"
122313 /translation="SGWSSVSRSLRPETSCDRPSSHLRNMVYNAELSPNSSSSG
ELUSVPEPRETFQALFGVYIDYTFQHKHLKVTNIDDMTETLYVKLPENMTRCDHL
PITCYLIGRSYGAIVAHADNATVGHVHDSIEPLVRGFOGLKDAVYFLNRHCGLF
LVDYLSACTSCHALFMPQFRCSLQDYGHWHDSIEPLVRGFOGLKDAVYFLNRHCGLF

```

```

HSDISPSNILDVFTDTMGMGRLVLDYGTASLHSDRNKMLDVLKSSGRQLRYLYCQ
RPFISIAKDYKPLCLLSKCYILRGAGHIPDFSACGPGAGTALRLDLQSLGYSLLYG
IMHLADSTHKIPYPNPDMGFDRSDPLYLQFAAPKVVLLEYLSQMNLNLDGLTSCG
ESPCVDVTAEHMSQFLQWCRSLKRFKESYFFNCRPFEPHPLGLVLAELLADDFGP
DGRRG"
misc_feature
complement(123506)
/note="DONOR: AAGGTGACT, possible donor"
complement(123941..124939)
/note="BGLF3 reading frame"
/codon_start=1
/db_xref="PID:g1334894"
/db_xref="SWISS-PROT:P03220"
/translation="WENAVKADPDPMLAREYGCLELALAEACODTPEQKLVETPL
KSFVLNLPDNRWHEARSGRVAEDDIYFSSALELPLRLPEEWFQGGQW
SRMPSPQEMGMCIEFEVDGMLRIALANKNDDEVIGQALQILNHSQTWTLVPEDP
LPMWALYGPGRSHCEERHCYVAAARGKRPILLPVTYPCANTAEALHLTRCVYA
LYLVDWKGEDIEADPPFDVSRNLKMAKOLCLLPQBPFCITRVCLLCLLHKONLNAQYK
RPVDYDPCLLITGEAERYMVDVAGNYREASTGTTVLFTYDLSGIVADMTYDE"
complement(124117)
/note="TATA: TATAAAA"
complement(124219)
/note="polyA signal: AATAAA"
124938..125915
/note="BGRF1 reading frame, homologous to RF 45 VZV and
spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
Northern blots in BGRF1 detect 2.7, 2.6, 2.1kb late and
1.9kb early RNAs. 2.6, 2.1kb RNAs very weak."
/codon_start=1
/db_xref="PID:g1334893"
/db_xref="SWISS-PROT:P03219"
/translation="MLYASQGRLENRLNALQODSTQGCIGALTPPSIMYTGAKSDR
WAHPLVGTIHASNLXCPMLRAYCRHYGPRPVASDESLLPMFGSPALHTPVQVMCL
LPEDRLQRLPPLNLEDEALTEFKTSVSARAILLEDPNFLERFVTSIASPLSG
QYKHPARLEAFQKOVHLHSFVFLISIKSLEITDMDIFQSAFGLSEMTLEKLHIFK
OKASVELLIPRHGKTWIVVAILLSILSNLSNVQIGYVAHQHVASAVTEIIDTLTKS
FDSKRVENKETSITFRHSGKISSTVMCATCFNKSIRGQTFHLLFVDEANFIKKEA
LPAILGFMLOKDAKLIIFISSVNSDAQTSFLYKLDQERLNVVSYVCQEHQPDFM
QDSMVSCFRLHIPSYITMDSNIRATTLFDGAFSTELMGDTSSLQSGLSRTRVD
DAINQLELCRVDTLNPVRVAGRLASSLYVVDPAYTNNTSAGTGIAAVTHDRADPNRV
IVLGLHEFLDLTGDAALQATCVVALVSVSTVTLHPLEVKVAVEGNSQASVAL
ASIIGESCPKCAFVHTKDKTSSLOWPHMLITNEKSAFERLIYAVNTASLSQSVTV
KHTFKKVVH"
complement(125113)
/note="TATA: TATAAT before BGLF3"
complement(125484)
/note="polyA signal: AATAAA, 3' end of 1.6kb late, 1.8kb
late, 3.0kb late and 3.7kb early RNAs"
complement(125863..126873)
/note="BGLF2 late reading frame, poor homology to RF 44
VZV"
/codon_start=1
/db_xref="PID:g1334895"
/db_xref="SWISS-PROT:P03221"
/translation="NASANSREQLRKRFLNKECLWLVDASTPQMKYITATTAVSAY
VYQIAGFPKPYMNVTLVLKPKKPTVTVYINGTLATVARPEVLTKVAGQPHSLT
LMYFGVSDVAGEAYPEIRGNPVVTCITDLTAHVFTTSTA VKTVEELODITPSEIIP
LGRGANWAEAGALYNEFYNMMDLCCPNWPTFPSLTHFINLLTRCDNGECVTCYAGA
HYNIIRGWTEDDSPGTSQCCLLPCTLANNDYVPIPTRHALLGLMFKPDPAPFVGL
RFPNPKMHPDMSRVLQGLANGKEVPCTAQPWTLLRFSDLXSRAMLYNCQVLKRQVHL

```

```

SY"
126277
/note="TATA: GATAAAA"
complement(126851..128374)
/note="BGLF1 late reading frame"
/codon_start=1
/db_xref="PID:g1334896"
/db_xref="SWISS-PROT:P03222"
/translation="MDVHIDNQVLSGLGTPLLVHLFVDPVTMAELCPNRPVNCGAWC
QTLFSDRTGLTRVCVFAARGMLPGRSHRGTEFTSPVYCDGELPELYNPFHVAALRF
YDEGLVGELOIYVLSLEFEGAKRALTDGHLIREASGVESAAAMQPIPIDGPPCGAG
IEHPVAAAQVHEHPKTYDLKQILLEITOEENRGEORLGHAGSPALCLGLRLRAGAETK
AAATSVSKHPALENPNINSRAGSGGGGAGTGGTVGSGALSVPVPSFKTRR
AIRSRALVIRGIAHIFSPHALYVVTYPELSAQGRHLRMTAVTHASPADDLAEVSTLGA
PERFRFLISVALRISAREKLAQMAQTAQOEIPVPIPTSYSRILYKNSDLIREAFT
VQTVSWSCWAKVATISNAPKTPDACLWIDSHPLYEEGASWANGKVIDSRPGLGVAA
SOLVALGTDGCHVHLATTSQGAFLVLPGGFVIKQALALTPEERYIILARGIRREQ"
complement(126929)
/note="TATA: TATTTAAA EEL8 late promoter before BGLF2,
gives 1.6kb late RNA"
complement(127237)
/note="TATA: TATAAAA, potential promoter for 1.8kb late
RNA"
128029
/note="polyA signal: AATAAA"
complement(128344..129021)
/note="BDLF4 early reading frame"
/codon_start=1
/db_xref="PID:g1334897"
/db_xref="SWISS-PROT:P03223"
/translation="MSDQGRLSLPRGEGTDENPNRHLCSYKLEFHLPLPESMASVF
ACWCGEYHVCDSGSECTLIHEGVGCVALTGNVGMGPHFQPALRPWTIRQDQDQD
KWEPEQVOGLVTVVNHLYHYFNENVTSGVSEALFDCGALRPHIPALVSFVRPCLL
MLFGASSEKVVVDVLSLYIHVLSIYSQKTVYGALLFKSTRNKRYDAVAKRMLWM
STLTKTC"
complement(128432)
/note="TATA: TATTAAA before BGLF1, potential promoter for
3.0kb late RNA"
128848
/note="BAM: Bam HI G/D"
complement(129054)
/note="TATA: TATTTC before BDLF4, potential promoter for
3.7kb early RNA"
129188..130348
/note="BDRF1 reading frame, homologous to RF 42 VZV and
spliced gene in HSV (Costa et al, 1985). Spliced from
BGRF1. Northern blots in BDRF1 detect 2.7, 2.6 kb late and
1.9kb early RNAs. Possibly also 1.8kb early RNA."
129374
/note="TATA: TATAAGC"
complement(129377)
/note="TATA: TATAAG"
129413
/note="DONOR: GTGGTAAGT possible donor"
130347
/note="polyA signal: AATAAA"
complement(130359)
/note="polyA signal: AATAAA, 3' end of 0.9kb late RNA,
2.3kb late RNA and 3.2kb late RNA"
complement(130362..131066)
/note="BDF3 late reading frame 9XNXT/S"
/codon_start=1
/db_xref="PID:g1334899"
/db_xref="SWISS-PROT:P03224"
/translation="MAHADRAGAVMAMILICETSLIWTSSTASAGNVTGTAVT
TPPSAGSPSTNQSTLTITTSAPITTTAILSTTTTFTTPTTPTTSTASTIN
TKTKYQAQITATEAGTGTSTGVNTVTRSSSTTSATRITNATTLATPLTSKGTSA
TKTAAELPTVPDERQPSLGYCLPWLTVFVGLTFLMLLILFAAGLMSAKNKPLDEAL
LTNAVTRDPSLYKGLV"
complement(131104)
/note="TATA: TATAAAA EEL4 late promoter before BDLF3,
gives 0.9kb late RNA"
complement(131127..132389)

```

```

/note="BDLF2 late reading frame"
/codon_start=1
/db_xref="PID:g1334900"
/db_xref="SWISS-PROT:P03225"
/translation="WDEQVAVHEGTVSHVISREDEGVVHERVVLASGERVEVFKAP
APRPREGRASTHEDFTVPAAPAAVGPPEPEPHPPMHIHANGGETKTNTDQONQOT
TRTNNAKAETAEHDDTMASGGQGRGAFISADLLSLSLGRMAAPASPMKSEVVC
GERMRFEKDYDEGEAEITAEPPRCGMSFVFIYCCYLAFALLAFGNPLPLPSFMP
VGAKVLKGRGRDGPVLSYGCPTNPFCKVTYLLIPAVINNVTYYPNNTDSHGSGHGE
AAALHVAALFESGCPNLQAVTNRRFTFNVRASGRVERLVODMQRVLASAVVAMHH
CHETYIYVDGVPGEFTIPTCFKDVLAFLRSLVNTCTAPLKTYSKGNWPNWGAAGGG
KRQCRVRLTDRSPAYLEEYVMVQ"
132266
/note="TATA: TATAAAA"
complement(132400..133305)
/note="BDLF1 late reading frame, poor homology to RF 41
VZV"
/codon_start=1
/db_xref="PID:g1334901"
/db_xref="SWISS-PROT:P25214"
/translation="MDLKVVVSLSRSLYTDIAKMOORIGCILPLASTHGTONVOGLG
LGQVYSLFVDPVSYMYNLSLCTLAVIDEVSVDLSLLIKIPGGTYAKNKYQPFQ
WGTGSLVMPVFGREHATVKLESNDVIFVPMUPTFAEEVLQKILLFNYSKRV
MQAPGNADMLDHMHGVSYSYLGHHYELALPEVPGPLGALLDNLSLFCIMVTLPR
ASMRLVGLIRHEHDDLNLFOEMVPEIARIDLDDLSVADDLSLRVMVMYTLQSLAS
LFNLGPRLATAAYSOETLATCWL"
complement(132476)
/promoter
/note="TATA: TATTAA before BDLF2, likely promoter for
2.3kb late RNA"
complement(133312)
/misc_feature
/note="polyA signal: AATAAA, 3' end of 4.5kb late RNA"
complement(133321..137466)
/note="BcLF1 late reading frame, homologous to RF 40 VZV
and major capsid protein of HSV"
/codon_start=1
/db_xref="PID:g1334902"
/db_xref="SWISS-PROT:P03226"
/translation="NASNEGVNRPFPYLTVDADLLSNLRQSAEGLPHSFLLVSKD
AREAGIKFVLGVYNTAIOYVRELETAALAVSCVNTKDLSDMTDGKIQFRLSVPTI
ANGDGRPSKQRTFIVKCHKHHISTEMELSHLDLILHSIPETPEVFAEVGAVKT
VKSALOFGDALLERGLINTVLSVKURHAPPMFLQTLADPTTERGFSKTVKSDLIAM
FKRLHGFVDFDLAEANGSGFSQVYRSRLSEMAVAVSGESVLKGVSTYTAGGEV
GGVFIYDNLVKLDTFLGEADNQTMDGSPSYAVFVGENLVTAIVSYGRVMTEFH
MARIYDPSKAGSTKSDLPAAVAGVDQPRVPIAAVIKLGNHAVAESLOKWNIDTF
SYPLNRMQVSYFYPVGLFMPNPKYTTSAIKMLDNPQQLPVEAWIWNKNLLAF
NLQNALKVLCHLPHTPAHTLASNAAPAPRRRTYSLOHRRPNHMVLIYDVEYD
NKYAPVTDIALKCLGPTDFLPSNIDLRLLELPHYDIYIGRDRGERARRAHRRL
MYGNLPTPLAPAFQARGQFETATSLAHVVDQAVIETQDTAYDTAPFAFFYVEA
MHHGEFEKVMVPLVSLCINTYWSRGLAFVNSFMKIFCRHLGNNAISKEAYSM
YRKIXGELLALQALMLAGSDVGSQVXCALLDPNLLPPVAYTDIFTHLLTVS
DRAPOIIGNEVYADT'LAAPQFIERVGNDEMAAQFVALYGYRVNGDHDHDFRLHGP
VYDEGHADVLEKIFYVFLPTCTNAHMCGLVDFOHVAQTLAYNGPASHHFTREDI
NLNLENTLRDLLEISDLRPTVGMIRDLASFWICTFTFRAVRVSYDNDVTQQLAPN
ADKREQTVLVNLVAFASERTAVTQCLFHAIPHFMYGDPRAVTAAMHQDVTFVM
QROKRAVAFANPOLFAEYREWSRPMGKYAAELPSLVSISGMTAHMKISPMAY
IAQAKLHPGVAMTVRDEILSENILFSSRSTSMFTGTPNVRREARVDAVTFEV
HEMASIDTGLSYSTMPARVAATIDMGITODPFSVPEAEFGNQVNDYIKAKV
GAQRNLTLDRLPTVLAGTNVNGAPGLCHQQAICEIIVPTADVAEYQKNSPRG
RAACVVCSCNYQVQAEGLLYDHSRDAAYETRSVNPWASGLSGLDIMNSSTRT
LVPLQSPCRFAEFNKEELLNRNGLYMNWYNSQRGGHPATSNTEQVPPVAGTDVF
WVF"
complement(133332)
/note="DONOR: RAGGTGGT possible donor"
complement(133352)
/promoter
/note="TATA: TATTAA before BDLF1"
complement(133386)
/promoter
/note="TATA: TATATAA"
135178
/misc_feature
/note="polyA signal: AATAAA"
135394
/promoter
/note="TATA: TATAAGT"
136624
/misc_feature

/note="polyA signal: AATAAA"
136868
/note="BAM: Bam HI D/c"
complement(137710)
/promoter
/note="TATA: TATTAAA EHLI promoter before BcLF1, gives
4.5kb late RNA"
137857
/note="TATA: CATAAAC"
137991..139718
/note="BcRF1 reading frame"
/codon_start=1
/db_xref="PID:g1632799"
/db_xref="SWISS-PROT:P25215"
/translation="MLAHINOVTRIPPCPPSGREARLKFHFESHTMLSPNNATL
REIRRAATNLIHHHLVDLTHASQPFFLTRSGALIRFVCCNCTLPNISIQQKA
GDPGDLEILQSGNGGPPASFPFSPSTGSLRCIVAASLLPEVSVGHQELSPURSR
SOGQTDVRSQDPARRVALLRRDGAPKDPPLGFGPHRGPAPKSDSESRDRA
PPPLDSSQASRLVPGVGFRLLVNTNRVINTKLVGSEPLVKRVNVPRLINNVF
ARKYVVKETAFTVSLFFTDGVCANLAIVNISTGYLSFLAMTSLRCFLPVAIYPA
VSNWASTATDLGLENSLVRENRSQVFTNFPSPVSCDGLNVSWFKAAATATISRKH
QOTLHOLIRITPIVTHREAKISRIKKNRFLTLELRNRSQOVHKKREGLLDCAS
LRLDPCINRIASEGLDFEFSKRSIAHKNRHECALLGHRHSANTVKLVNDRKTKTA
ILGRNANLTRCKHQVNRQSPIFLTLLRHTRRLGRASVYKREITLILLAHLRKKTA
PIHCRAQV"
138019
/note="BAM: Bam HI c/b"
139352
/note="BAM: Bam HI b/t"
139642..140916
/note="BTRF1 reading frame. Northern blots detect 0.95
late and 3.8kb early RNA; Protein sequence is in conflict
with the conceptual translation"
/codon_start=1
/db_xref="PID:e25011"
/db_xref="PID:g1334904"
/db_xref="SWISS-PROT:P30119"
/translation="NERLPESWPTCAKHQPPSTAVMLKCKOPGARFIHGAVHLPSSQI
VFHTHSTLASALGPGENPIPALFRASGLNVRESLPMTNRPAPISLARLIAPN
PYLHGQTVGTMQDNGPVLFAFARVIEKSPESNIKASSQLMAEPEVYQLRCHLSVAL
ASEHPAFSVEVSKRVYDEGNARTRTIRDLIEPVTLSQLSPSTKSLKKAPEPP
PEPQATDATPYARIFYDIGRVQKLGNAAPAAQVSNLIANRHSRLVNPDLPL
LQHLVLKWLKSLNLENIVODEAIFTSPDSTISEATKAPEKLEVOAKNVENTVF
CLNSICSTLPDVPVDPVNNPDISLAKSKYFLMFPSPGTIMRNVRFATIVRLCQGA
ELGTNAQFLGKVKVKKETGMVTLVKLYLLRI"
complement(140902)
/note="polyA signal: AATAAA, 3' end of 2.5kb late RNA"
complement(140916..143036)
/note="BXLF2 late reading frame, encodes gp85; homologous
to RF 37 VZV and glycoprotein H of HSV (gpIII of VZV)"
/codon_start=1
/db_xref="PID:g1334905"
/db_xref="SWISS-PROT:P03231"
/translation="MOLLICVFLVLLHVEGAASLSEVKLHLDIEGHASHYTIPTWELM
AKVGLPEALWREANVTEDLASMLNRKLYIKTSGTILGIALAEVDPVPAVEGSMQV
DASKVHPGVISGLNSPACMLAPLEKQLFYITGMLPNTRPSHYVYQLRCHLSVAL
STNGDKFOYTGAMTSKFLMGYKRVTEKGDHVLSTLVGKTKDLPDLRGPSYPSIT-
AGQSDYSLVITFTFHYANFNHYFVNLKDKHFSRAVMTAASYARYVLOKLVLEKKG
GCREPDELTLTITMFSVAVFFKVGHAVGETNGCVDLRWLAKSFFELVTKDILGI
IGGATVGMQSYGLERLAAMATMYMEELHGLTTERQEYALRLATVGYPAKVYSGL
TCGATSVLLSAYNRHPLFQPLHTVMRETLFGSHVVLRELRLNLTQGPNLALYQLLS
TALCSALBIGVLELRGLATESGLFSPCVLSLRFLDTRDKLLSMAPOEATLDOQAVSN
AVDGFGLSLRLERDADHLPAYKCDVRLDKVMTIPLINVTFIISDREVRGSALY
EASTYILSSSLFSPVIMNKSQGAAGEPROIPKIQNFTRQKSCICGCFALLSYDE
REGLETTITTSQEVNSILSNFYDNLHVHLLLTNGTNGVMEAGLYEERAHVVL
AIIIFYIAFALGIFLVHVMFFL"
140970
/note="polyA signal: AATAAA"
complement(141286)
/misc_feature
/note="polyA signal: AATAAA"
142589
/promoter
/note="TATA: GATAAAA"
142740
/misc_feature

```

```

/notes="BAM: Bam H1 T/X"
complement(143038)..144861)
/notes="BXLf1 early reading frame, thymidine kinase
(Littler et al, 1986). Weak homology to RF 36 VZV and HSV
thymidine kinase. 4.0kb early RNA presumably encodes the
TK. Also a 2.2kb late RNA here."
/codon_start=1
/db_xref="PID:g1334907"
/db_xref="SWISS-PROT:P03177"
/translation="MAGPFGKEAGPPGGWGRKQCEDESPENHERHENFYAEIDDFAPSVL
TPGSDSGEDDGLQVPTHPPLMAPTGLSERVPCPTQAAVTISCTNGPSGRH
TSCPTLPRGAQPPAPAKOKTAPTPKPSRECGSKTPDPESWFKTSCTEGGADST
SRSFYQKGFEGLAGLDDKSDCESEDESNFRPPSHSALKQKNGKGPGLGFH
LAHGREFSKSHKAQKRLUGSVNMVNLDDAODTQKAQAQRKESMRVPIVTHLTN
HVPYTKPACSLFLEAGPVGKTMNLHLKAVFGDLTIIVPEPMRYTWTHYNAIKAMH
KNVTRARIGREDTSAEVLACQMKFTTFFRVLASRRKSLVTESGARSVAPLDCWLLHD
RHLSAVVFLPMLRLSOLLSSDFIOVLATFTADPGDTIVMKLVNENMRRLKRG
RKHESGLDAGLYKSVNDAYHAYVCAMLLTOYFAPEDIYKVCAGLTTITTVCHQSHPTI
IRSGVAELKXNSIFSVELKVIQPPRADAVLLEVLATRTILAYLQFLVLDLSEFQDD
LPGCWTEIYMALKNPARKSRQFFDWAGUSKVIISDFERNRD"
complement(143310)
/notes="TATA: TATAAGA ECL2 late promoter before BXLf2,
gives 2.5kb late RNA"
/notes="polyA signal: AATAAA"
143608
144791
misc_feature
misc_feature
promoter
/notes="ACCEPT: TCTTCGTTTTCAGG poss. acceptor before
BXRf1"
144860..145606
/notes="BXRf1 late reading frame, homologous to RF 35 VZV.
Basic (core) protein."
/codon_start=1
/db_xref="PID:g1334906"
/db_xref="SWISS-PROT:P03232"
/translation="MDPTRGICLALSTHDLAKFHSPPARKAAGKRAHLRCYSKLLSLK
SWEQASLSPGPTFTDFDLRFVFTLGRRIADCVVVALQYPRCYIVEKTKMSNT
ANQSVTKRAORLETCDCANFLRSCPPVLGSGGLEVALAALVFNKSLTKLQV
EPFALGQKLTPTSTGLNLSRWQDGLARLDRPRPTAQGHRPRTHVGPKPQLTA
RVPRSRAGRAGGRKGQVGAVGQVCPGAQK"
144862
misc_feature
144945
misc_feature
/notes="DONOR: CAGGTAAAC possible donor at 3' BXRf1"
complement(145135)
/notes="TATA: TATAACA before BXLf1"
145302
promoter
/notes="TATA: TATTTAA before BVRf1, potential promoter for
1.9kb early RNA"
145416..147128
/notes="BVRf1 early reading frame, homologous to RF 34 VZV"
/codon_start=1
/db_xref="PID:g1334908"
/db_xref="SWISS-PROT:P03233"
/translation="MALSGHVLIDPARLPDRTGPELMPWASLRSLRVSPALELAER
EABRSRERWCAQVLNRLRLVELDGMRDHLARAEIRQDLDVAVFSDGLESMQ
VRSPTGSRGAPAPPSPAPQFTRLTGNQYAVSISPTDPLVAVGSLAGTLGNLY
GNINQVPSFGVPTMTSANAMORRVFKQLGRNFTNSVSLKMTVEVAVLEGTQD
DFFSDVRLPDLOAALILSVAYLLQGGSSHQORLPASRELELPGESLEKIAD
KAKSPGFMILTSNKEAROSIAPLNQQAAYPGTFADNKIYNLNVGAGLLPTAAL
NYPGAARDRLDIYRIANGIFGEDVPFSSHOWNRVGLAALMLVITTCETANLA
EAATRHLSSLPLQAMQRRKPAMASAGMPGAYPVQTLFRHGELFRFTWAVYRPTVA
ADQASISLSPGLVLALELKLMDQAPSHYAINLTQKQFTDLFEINOKLLFHPA
AMLAARTQLRLAFEDGVGVALGRSPMLAAREILLERQFSASDDYDRFLTLGLVASP
VAPS"
complement(146926)
/notes="polyA signal: AATAAA"
147167
/notes="DONOR: AAGTAAAT possible donor"
147170
/notes="polyA signal: AATAAA, 3' end of 2.4kb late and
1.9kb early RNAs"
147721
promoter
/notes="TATA: TATTAT before BVRf2, potential promoter for

```

```

2.1kb early RNA"
147927..149744
/notes="BVRf2 early reading frame, N-terminus homologous to
RF 33 VZV"
/codon_start=1
/db_xref="PID:g1334909"
/db_xref="SWISS-PROT:P03234"
/translation="MVQAPSVYCGFVERDAPPDAPKDACLHLDPLTVKSQLKPLPL
PVEHLAPAPGVSGFLYQSRAGLSAASITSGDFLSLDSIYHOCIDIAQSORLPLPRE
TKVALLHAWPLSLASLHPDIPOTTADGGKLSFDFHVSICALGRRRTTAYGTDLA
WVLFHSDLEPSIAQLENDANAKRSGCCPDHPLTLKLIKAIDAGFLNRVETIL
RODGVANIAPRAESYLKASDAPLQKPKALQSPSPASTDPAIMLSGNAGATACGGS
AAAGODLISVRNTFTMTLQTNLDNKKPPRQTPPLAAPLPFSHAOTATASYGPAG
AAVAGGYFTSPGGYIYAGAGDGFGLAMDADHTYHPHPHPAYFGLPGLFGPPPY
PPYCGSHRADYVPAPSRNKRKRDEDEEGGLFCGEDATLYRKDLAGLSKSNVEL
HTQALARRLETLSYGTGVCYQPGQPCYTHSGYGVGFQSHQSYEVPRVYHPHPPTS
HQQAQOPPPGCTOAPEAHCVAESTIPEAGAGNAGSGPREDNPQOPTTEGHRGKLV
QASASGVAQSKPTTPRAKSVSAHLKSFCEELLNKRVA"
148007
/notes="BAM: Bam H1 V/d"
148620
/notes="TATA: TATTAA ECR1 late promoter before Bdrf1,
gives 1.2kb late RNA"
148707..149744
/notes="Bdrf1 reading frame; this is the C terminus of
BVRf2"
/codon_start=1
/db_xref="PID:g1334910"
/db_xref="SPTRMBL:Q66541"
/translation="MLSGNAGEATACGGSAAAGODLISVRNTFTMTLQTNLDNKKPP
RQTPPLAAPLPFSHAOTATASYGPAGVAGYFTSPGGYIYAGAGDGFGL
AMDADHTYHPHPPPAYFGLPGLFGPPPYGSHLRADYVPAPSRNKRKRDEE
DEEGGLFCGEDATLYRKDLAGLSKSNVELQALARRLETLSYGTGVCYQPGQPC
YTHSGYGVGFQSHQSYEVPRVYHPHPPTSHTQQAQOPPPGCTOAPEAHCVAESTI
PEAGAGNAGSGPREDNPQOPTTEGHRGKLVQASASGVAQSKPTTPRAKSVSAHLKSI
FCEELLNKRVA"
149115
misc_feature
149727
/notes="BAM: Bam H1 d/i"
149727
/notes="polyA signal: AATAAA, 3' end of 2.1kb early and
1.2kb late RNAs"
complement(149758)
/notes="polyA signal: AATAAA, 3' end of 1.0kb late, 1.5kb
late and 1.8kb late RNAs"
complement(149779)..150525)
/notes="BILF2 late reading frame 11xNXT/S"
/codon_start=1
/db_xref="PID:g1334911"
/db_xref="SWISS-PROT:P03218"
/translation="MTHLVLLCCCVSGVSCAFFSDLVKFENVTAHAGARVNLTCVSPS
NESVRIELGRGYTPGDQLPLAVATNNGTHITNGVYNSLTLEWVNDSTSVSLII
PNVTLAHAGYTCNVTLRNCVSVAGVHCNYSAGEDDOYHANRILTQRMHLITVPAIT
IAPTLVSHSTISHRHPRVPSRPHKPVTLGLTFPFDIPWRPKTTVHWALLLITCA
VAPVLLIIISCLGLAGNRRKRWIPL"
complement(150571)
/notes="TATA: TATTAG before BILF2. Potential promoter for
1.0kb late RNA."
151236..151618
/notes="repetitive sequence 3x"
151767
/notes="polyA signal: AATAAA"
complement(151780)
/notes="TATA: CATATAA"
152012..152013
/notes="DEL: B95-8 deletion with respect to Raji"
complement(152161)..153099)
/notes="BILF1 reading frame, membrane protein, 3xNXS /T"
/codon_start=1
/db_xref="PID:g1334912"
/db_xref="SWISS-PROT:P03208"
/translation="MLSTMAPGSTVGLVANMTSVNATEDACTKSYSAFLSGMTSLLL
VLLILLIAGILFIIFVRKLVHRMDVMIALLIELLVLGKMQIEFSSTGLCLTQ
MMFLGLMCSVMTHLGMALFKTLALFSRTPKRTSHRNVCVLYLGMVFCVLILLIILIT

```



	/note="BARE1 reading frame" /codon_start=1 /db_xref="PID:gl334917" /db_xref="SWISS-PROT:P03228" /translation="MARFIAQLLLASCVAAGQAVTAFLGERVTLTYSWRRVSI GP EVSFKLPGCEQVILGMHHDVIFIEWPFGDFDIHRSANTFLLVVTANISHDGN LCRMKLEPTEVTRKHLSVAPKPLTSVHRSRQPDFSVLTICTVNAFPHPHVQWIM PEGVEPTAANGVMKEKDGSLSVAVDLSLPKPWHLPVTGCKNDKEEAHGVIYVSG LS"
promoter	complement(165713) /note="TATA: TATAAG before BALF1"
misc_feature	166165 /note="polyA signal: AATAAA 3' end of 0.8kb early RNA. Also 1kb late RNA in this region."
promoter	166469..166475 /note="TATA: TTATTTT"
mRNA	166498..166916 /note="Exon 1 of terminal protein RNA"
misc_feature	166561..166563 /note="Likely initiator met of terminal protein"
misc_feature	166614 /note="BAM: Bam HI A/Nhet"
misc_feature	complement(166946) /note="polyA signal: AATAAA These 2 polyA sites are 3' end of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"
misc_feature	complement(166950) /note="polyA signal: AATAAA"
CDS	complement(166998..167303) /note="BNLF2b reading frame; Protein sequence is in conflict with the conceptual translation"
	/codon_start=1 /db_xref="PID:e25001" /db_xref="PID:gl334919" /db_xref="SPTREMBL:O04362" /translation="EATMRGRLAGYATILRSFRMSKRSKNKAKKERVPEYEDPP TPMTSRLIRNALGGVVRPDAEDCQRFHPLEPALGVSTKFNFDLLSLRCELGWCG" complement(167304..167486) /note="BNLF2a reading frame"
CDS	/codon_start=1 /db_xref="PID:gl334920" /db_xref="SPTREMBL:O04361" /translation="MVHVLERALLEQQSSAGLPGSSTETRPSHPCPDPPVSRLLRL LVLCVLFGLLCLLLI" 167320
promoter	/note="TATA: CATAAAA"
promoter	complement(167525) /note="TATA: TATAAA EDL2 early promoter before BNLF2a,b. Gives 0.8kb RNA"
mRNA	complement(168163..168965) /note="BNLF1 coding part of exon c of latent membrane protein"
repeat_region	168399..168574 /note="5 x 33bp repeats"
intron	complement(168966..169041) /note="intervening sequence in BNLF1"
misc_feature	complement(168966) /note="ACCEPT: TTTTTCACGCAGT possible acceptor"
misc_feature	complement(169041) /note="DONOR: TTGCTAAGA donor"
mRNA	complement(169042..169128) /note="BNLF1 exon b of latent membrane protein mRNA"
intron	complement(169129..169206) /note="intervening sequence in BNLF1"
misc_feature	complement(169129) /note="ACCEPT: TCCTCTTTCCCAAGT acceptor"
promoter	complement(169201) /note="TATA: TATTACA EDL1a late promoter, gives 2.5kb late RNA"
misc_feature	complement(169206) /note="DONOR: TGAGTAAGT donor"
mRNA	complement(169207..169474) /note="BNLF1 exon a of latent membrane protein mRNA"
promoter	complement(169546)

[illegible]

RESULT 12			
HS4B958RAJ/c			
LOCUS	HS4B958RAJ 184113 bp	DNA	VRL 12-APR-1996
DEFINITION	Epstein-Barr virus, artifactual joining of B95-8 complete genome and the sequences from Raji of the large deletion found in B95-8		
ACCESSION	M80517	M75989	





```
QY 1556 CCTCCACCGTGGTCCACCTGCAACGGCCACCGCAACTGAGCCCGACCTGCCACGTACA 1615
Db 927 GCGCTCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCT 868
QY 1616 ACCGCAATTCCTACTCGCGCCCGCCCGCTGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1675
Db 867 CTTGCGCCCTCTGCGCCCTCTGCTGCTGCGCCCTCTGCTGCGCCCTCTGCTGCGCCCTCTGCT 808
QY 1676 ---CCTGAATCCCAACCACTGTACAGTACACCTACTGCTGCTGCTGCTGCTGCTGCTGCT 1732
Db 807 CTTGCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 748
QY 1733 ACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1792
Db 747 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 688
QY 1793 GCTGCGCCCGCCCGCTGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC 1852
Db 687 CTTGCGCCCTCTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 628
QY 1853 ACTGTCATATCACCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1912
Db 627 CTTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 568
QY 1913 GTCACTGTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1972
Db 567 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 508
QY 1973 ACTGCAACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2032
Db 507 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 448
QY 2033 GTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2092
Db 447 GCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 388
QY 2093 ACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2152
Db 387 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 328
QY 2153 GACCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2166
Db 327 CTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 314
```

## RESULT 14

```
D88734
LOCUS D88734 3173 bp DNA VRL 19-MAY-1997
DEFINITION Equine herpesvirus 1 DNA for membrane glycoprotein, complete cds.
ACCESSION D88734
NID 92114322
KEYWORDS membrane glycoprotein.
SOURCE Equine herpesvirus 1 (strain: BK343, isolate: 3F clone) DNA.
ORGANISM Viruses; dsDNA viruses, no RNA stage; Herpesviridae;
Alphaherpesvirinae; Varicellovirus.
REFERENCE 1 (bases 1 to 3173)
AUTHORS Kirisawa, R., Kobayashi, T., Kawakami, Y. and Iwai, H.
TILE Direct Submission
JOURNAL Submitted (31-OCT-1996) to the DBJ/EMBL/GenBank databases. Rikio
Kirisawa, Rakuno Gakuen University, Faculty of Veterinary Medicine,
Department of Veterinary Microbiology; Bunkyo-dai-Midorimachi 582,
Ebetsu, Hokkaido 069, Japan (E-mail: r.kirisawa@ca2.so-net.or.jp,
Tel: +81-11-386-1112, Fax: +81-11-387-5890)
REFERENCE 2 (sites)
AUTHORS Kirisawa, R., Kobayashi, T., Kawakami, Y. and Iwai, H.
TILE Nucleotide sequences of open reading frames 1, 24 and 71 of an
attenuated equine herpesvirus-1
JOURNAL J. Equine Sci. 7, 79-87 (1996)
FEATURES Location/Qualifiers
source 1..3173
/organism="Equine herpesvirus 1"
/isolate="3F clone"
```

```
/strain="BK343"
/db_xref="taxon:10326"
118..2721
/gene="ORF71"
118..2721
/gene="ORF71"
/codon_start=1
/product="membrane glycoprotein"
/db_xref="PID:d1020846"
/db_xref="PID:g2114323"
/translation="MGFIYARKLLKLVNAVSIYAIGSTTTTETTTSSSTSGSGOSTSS
GTNSSSPSTSPSTSSSTSPSTSHSTSSPSTSTQSSSTAASTSSASTSSASTSTSTPT
STSTSTSTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
TPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
TTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA
SGSTSTGASTSTPSASTATSTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
TTTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
TTTPTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST
STEALSTPDGNTSGNTSPSPSTPSFADTQTPDNGVSTQHTTINDHTTANAKHA
GYHRGAGRGRIPOGSHTPHPDRLTPSPDDTDDTNDHNPGRNNSIEIVPOLPPD
RPIELGVATLRKNFMEASCTVETNSGLAIFWKIGNASVDAFNRGTTHTRLMRNGVPV
YALVSTLRVPLNVIPLTKITCAACPTNLVAGDGLNSCTTKSTTTPCPCQOQTHIF
FSAGKDRVCITSELVSQPTITWSGDRDLNDGFSQIWIQIGVCGILRSEVRHHR
TTMFGSTSKDYLCEVSASDSKTSYDKVLPNAHSTSNFALVAATTLIVTILCLLCCLY
CMLTRPRASVY"
repeat_region 667..861
repeat_region 892..1101
polyA_signal 2802..2807
BASE COUNT 857 a 1068 c 543 g 605 t
ORIGIN
```

Query Match 5.3%; Score 149.4; DB 17; Length 3173;

Best Local Similarity 53.0%; Pred. No. 7.2e-16;

Matches 349; Conservative 0; Mismatches 301; Indels 9; Gaps 1;

QY 1480 CACATCGACTGCTGCCCGACAGCTGCTCCAAAGCGCTGCCCAAGCGCTGCCCAAGCGC 1539

Db 459 CGCATCTACAAGCACCACCAACACACCGCGCTCCCAACAGCGCGCTTACACCAAC 518

QY 1540 TGCCCCAACTCGCGACCTCCACCGTGTGCCACCTGCAAGCGCACCGCACTGCGAGC 1599

Db 519 AGCTGTACTACGCGCGTCTACATCAGCGAAACACCACTACTACGCTACCCACCTGT 578

QY 1600 CCCAGTCCCACTCAACCGCAATTCCTACTCGGCGCCCGCCCGCTGCCCGCCCGCAGC 1659

Db 579 CTCACCCCAACCACTACGCTTACCTCCACAACTACTACGCTACCCACCTGT 638

QY 1660 TCCTACTACTGCTGCCCTGAATCCCAACCACTGTACAGTACCACCTACTGCTGTCC 1719

Db 639 TCCAACAACCGCTTCTACAACACTGATACGACCAACGACGACGACGACGACGACGAC 698

QY 1720 CACCGCAGCCCTACTACTGCTGCTCCCTGAAATCCCAATCACTGTACATCAGCGCTAC 1779

Db 699 GACCAACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 758

QY 1780 CGCTGCCCCCGCTGCCCGCTGCCCGCTGCCCGCTGCCCGCTGCCCGCTGCCCGCTGCC 1839

Db 759 GACCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 818

QY 1840 AGAATTCCTCACTACTGTCTACCTACCTACTGCTGCCCGCCCTACTACGCGACCTGTC 1899

Db 819 GACCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 878

QY 1900 CCCCAACACCACTACTGTACCACTGCTGCCCGCTACTACCGGACGACCTGTCTCC 1959

Db 879 TACTTCTCTGCAACACCGCGGTACCAACCGCGGTACCAACCGCGGTACCAACCGCGGTAC 938

QY 1960 CAACACACAGTACTGACCACTGCGGAGCGGCTTACTACCGGACGACCTGTCCCA 2019

Db 939 CACCGCG-----GCTACCAACCGCGGTACCAACCGCGGTACCAACCGCGGTACCA 989



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein --protein search, using sw model

Run on: July 18, 1999, 01:02:57 ; Search time 19.54 Seconds  
(without alignments)  
1082.403 Million cell updates/sec

Title: US-09-294-663-3  
Perfect score: 4328  
Sequence: 1 MIKTLFLTALGLVAARPEV.....VFIPWNKLEIDIRQALNPEL 788

Scoring table: BLOSUM62

Searched: 74019 seqs, 26840295 residues

Database : SwissProt\_36.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	636	14.7	5179	MUC2_HUMAN	Q02817 homo sapien
2	433	10.0	797	VLGX_HSVB	P28968 equine herp
3	432	10.0	662	MUC1_XENLA	Q05049 xenopus lae
4	431	10.0	865	CPN_DROME	Q02910 drosophila
5	397	9.2	1257	MUC1_HUMAN	P15941 h mucin 1 p
6	393.5	9.1	1365	AMYH_YEAST	P08640 saccharomyc
7	353.5	8.2	1664	SLP1_CLOTH	Q08852 clostridium
8	351	8.1	474	VTP3_TTV1V	P19275 thermoprote
9	350.5	8.1	279	Y091_NPVOP	O10341 orgyia pseu
10	349.5	8.1	822	ANP_NOTCO	P24856 notothenia
11	331.5	7.7	400	MUAI_XENLA	P10667 xenopus lae
12	330.5	7.6	670	VG50_HSVII	Q00130 ictalurid h
13	309	7.1	907	VG3_P_EBV	P03200 epstein-bar
14	309	7.1	1161	YJ9P_YEAST	P47179 saccharomyc
15	308	7.1	860	CHT2_COCIM	P54197 coccidioid
16	307.5	7.1	307	SGS3_DROME	P02840 drosophila
17	306	7.1	267	EXTN_MAIZE	P14918 zea mays (m
18	301.5	7.0	886	VG3_EBV8	Q07284 epstein-bar
19	297	6.9	620	EXTN_TOBAC	P13983 nicotiana t
20	294.5	6.8	3149	TEGU_EBV	P03186 epstein-bar
21	292	6.7	283	EXTN_SORVU	P24152 sorghum vul
22	286	6.6	634	HWPI_CANAL	P46593 candida alb
23	281	6.5	1794	YAV1_CHPO	O10172 schizosacch
24	279	6.4	2476	ZAN_PIG	Q28983 sus scrofa
25	271	6.3	1251	IYQ3_CABEL	Q09550 caenorhabdi
26	270.5	6.2	263	SGS3_DROYA	P13728 drosophila
27	269.5	6.2	1162	TCNA_TRYCR	P23253 trypanosoma
28	269.5	6.2	699	VGLG_HSV2H	P13290 herpes simp
29	269	6.2	338	SED1_YEAST	Q01589 saccharomyc
30	263	6.1	600	SP96_DICDI	P14328 dictyosteli
31	260	6.0	467	RP61_CRIGR	P11414 cricetus
32	260	6.0	1970	RPB1_HUMAN	P24928 homo sapien
33	260	6.0	1970	RPB1_MOUSE	P08775 mus musculu
34	259.5	6.0	3421	TEGU_HSVB	P28955 equine herp
35	257.5	5.9	721	YV21_MYCTU	O10690 mycobacteri
36	255.5	5.9	901	AI80_MOUSE	Q61548 mus musculu
37	253	5.8	1083	T2D3_HUMAN	O00268 homo sapien
38	249	5.8	1742	GUNA_CALSA	P22534 caldocellum
39	248	5.7	439	XP2_XENLA	P17437 xenopus lae
40	247.5	5.7	915	AI80_RAT	Q05140 rattus norv
41	244	5.6	796	YS8A_CABEL	Q09625 caenorhabdi
42	243	5.6	442	SDC3_RAT	P33671 rattus norv
43	242	5.6	1183	DRPL_RAT	P54258 rattus norv

#### ALIGNMENTS

```
RESULT 1
MUC2_HUMAN
ID MUC2_HUMAN STANDARD; PRT; 5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
GN MUC2 OR SMUC.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;
OC EUHERIA; PRIMATES.
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE; 94132002.
RA GUM J.R. JR., HICKS J.W., TORIBARA N.W., SIDDIKI B., KIM Y.S.;
RL J. BIOL. CHEM. 269:2440-2446(1994).
RN [2]
RP SEQUENCE OF 626-1895 AND 4196-5179 FROM N.A.
RC TISSUE=COLON;
RX MEDLINE; 93016075.
RA GUM J.R. JR., HICKS J.W., TORIBARA N.W., ROTHE E.-M., LAGACE R.E.,
RL KIM Y.S.;
RN [3]
RP SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE; 91358717.
RA TORIBARA N.W., GUM J.R. JR., CULHANE P.J., LAGACE R.E., HICKS J.W.,
RL PETERSEN G.M., KIM Y.S.;
CC J. CLIN. INVEST. 88:1005-1013(1991).
CC -!- FUNCTION: COATS THE EPITHELIA OF THE INTESTINES, AIRWAYS, AND
OTHER MUCUS MEMBRANE-CONTAINING ORGANS.
CC -!- SUBUNIT: MULTIMERIC.
CC -!- SUBCELLULAR LOCATION: SECRETED.
CC -!- TISSUE SPECIFICITY: COLON, SMALL INTESTINE, COLONIC TUMORS,
BRONCHUS, CERVIX AND GALL BLADDER.
CC -!- PTM: ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR
INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND
VARIES AMONG DIFFERENT ALLELES.
CC -!- SIMILARITY: THE N-TERMINAL DOMAIN SHOWS SOME SIMILARITY TO THAT
OF SILKWORM HEMOCYTIN.
CC -!- SIMILARITY: CONTAINS 2 VFPC DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK).
EMBL; L21998; G454154; -
EMBL; M74027; G188864; -
EMBL; M94131; G186396; -
EMBL; M94132; G186398; -
HSSP; P15445; 1PSH.
MIM; 158370; -
PROSITE; PS00022; EGF_1; UNKNOWN_1.
PROSITE; PS01185; CTCK_1; 1.
PROSITE; PS01225; CTCK_2; 1.
PROSITE; PS01208; VFPC; 2.
GLYCOPROTEIN; REPEAT; POLYMORPHISM; SIGNAL.
SIGNAL 1 20
CHAIN 21 5179
DOMAIN 1401 1747
REPEAT 1401 1416
REPEAT 1417 1432
REPEAT 1433 1448
REPEAT 1449 1464
REPEAT 1465 1471
REPEAT 1472 1478
REPEAT 1479 1494
```

```
FT REPEAT 1495 1517 7B.
FT REPEAT 1518 1533 8A.
FT REPEAT 1534 1556 8B.
FT REPEAT 1557 1572 9A.
FT REPEAT 1573 1596 9B.
FT REPEAT 1597 1612 10A.
FT REPEAT 1613 1635 10B.
FT REPEAT 1636 1651 11A.
FT REPEAT 1652 1675 11B.
FT REPEAT 1676 1683 12.
FT REPEAT 1684 1699 13.
FT REPEAT 1700 1715 14.
FT REPEAT 1716 1731 15.
FT REPEAT 1732 1747 16.
FT DOMAIN 4815 4886 VMFC 1.
FT DOMAIN 4924 4991 VMFC 2.
FT DOMAIN 5075 5160 CUCK.
FT DISULFID 5075 5122 BY SIMILARITY.
FT DISULFID 5089 5136 BY SIMILARITY.
FT DISULFID 5098 5152 BY SIMILARITY.
FT DISULFID 5102 5154 BY SIMILARITY.
FT DISULFID ? 5159 BY SIMILARITY.
FT CARBOHYD 163 163 POTENTIAL.
FT CARBOHYD 423 423 POTENTIAL.
FT CARBOHYD 670 670 POTENTIAL.
FT CARBOHYD 770 770 POTENTIAL.
FT CARBOHYD 894 894 POTENTIAL.
FT CARBOHYD 1139 1139 POTENTIAL.
FT CARBOHYD 1154 1154 POTENTIAL.
FT CARBOHYD 1215 1215 POTENTIAL.
FT CARBOHYD 1230 1230 POTENTIAL.
FT CARBOHYD 1246 1246 POTENTIAL.
FT CARBOHYD 1787 1787 POTENTIAL.
FT CARBOHYD 1820 1820 POTENTIAL.
FT CARBOHYD 4339 4339 POTENTIAL.
FT CARBOHYD 4351 4351 POTENTIAL.
FT CARBOHYD 4362 4362 POTENTIAL.
FT CARBOHYD 4373 4373 POTENTIAL.
FT CARBOHYD 4422 4422 POTENTIAL.
FT CARBOHYD 4438 4438 POTENTIAL.
FT CARBOHYD 4481 4481 POTENTIAL.
FT CARBOHYD 4888 4888 POTENTIAL.
FT CARBOHYD 4955 4955 POTENTIAL.
FT CARBOHYD 4970 4970 POTENTIAL.
FT CARBOHYD 5019 5019 POTENTIAL.
FT CARBOHYD 5038 5038 POTENTIAL.
FT CARBOHYD 5069 5069 POTENTIAL.
FT CONFLICT 1351 1351 H > L (IN REF. 3).
FT CONFLICT 1412 1412 T > S (IN REF. 3).
FT CONFLICT 1449 1449 L > P (IN REF. 3).
FT CONFLICT 1504 1504 M > T (IN REF. 3).
FT CONFLICT 4192 4192 G > S (IN REF. 2).
FT CONFLICT 5179 5179 AA; 540295 MW; ECD04582 CRC32;
SQ SEQUENCE 5179 AA; 540295 MW; ECD04582 CRC32;

Query Match 14.7%; Score 636; DB 1; Length 5179;
Best Local Similarity 25.6%; Pred. No. 1.2e-20;
Matches 201; Conservative 41; Mismatches 213; Indels 330; Gaps 24;

QY 29 LHEPHP-----DCPAAEQHLLPHEVDCTKF-----YCEVGLKFIAPRD 68
DB 1309 INEDHPSSGSDGDPEDFGVCGAPEDIKRSVKDPHLSLEQHGKQVQCDVSVGFIKNE 1368
QY 69 CAPGT-----BFFSAQTCVHAALAGC-TLPGPAAETQAPATTQATTQATTQ 120
DB 1369 DQFGNGPFLCYDKIRVNC--WPMDCITTPSPPTTTPSPPTT---TTLPPTTPS 1423
QY 121 APTTTQAPTTTQAPTTQAP-----TTTQAPTTQAP-----TTTQAPTTTQAPTTTQ 173
```

```
DB 1424 PPTT-----TTTTPTTTPSPPTTTTTLPTLPTTTPSPPTTTPPTTTPSPPTTTPS 1478
QY 174 APTTQAP-----TTTQAPTTTQAPTTTITCAATTQAPATT--PAATTQAPATTQAPATT 227
DB 1479 PPTTTPSPPTTTPPTTTPPTTTPSP--PMTPTTPPASTTTLPTTTPSPPTTTPPTTTPPTT 1537
QY 228 PGVPAETSAVMPPICELLPNCGPADFIDHLLIPHDKYNLFYQCSNGYTFEQRCEPLY 287
DB 1538 TPSP-PTTTPITPP----- 1551
QY 288 FNYVORCDSPANVECDGEISPAFPVTEGNEDEIDIGLLONGCPANFEIDWLLPHGNR 347
DB 1551 ----- 1551
QY 348 CDKYYQCVHGNLVERRCGAGTHSFELQOCDHIELVGTLPGESEEDVDDEACTGWYC 407
DB 1551 ----- 1551
QY 408 PTEPIEWELPNCGPADFSIDHLLPHESDCGOYLOCVHGQTIARPCGNLHFSPTQSC 467
DB 1551 ----- 1551
QY 468 SPVTACQVFECDSDNQCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 518
DB 1551 -----TSTTTLPTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTITTT 1594
QY 519 TPTATAAPVPPTTAIPTAPTAAPTAAPTAAPESPTTVTPETAAPTAAPTAAPTAAPTAAP 578
DB 1595 TPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1654
QY 579 TVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 637
DB 1655 TTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1711
QY 638 PTTAAPAPNTTVTPETAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 694
DB 1711 -TTTTPSSTTTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1769
QY 695 -----PNDPINCVECNVLPWAHADCDKYWCDDGNVL-VVCSGLQFNPTTKTCDEFA 748
DB 1770 FSPFSTTTTTPCPLCNWGTGWDGKPNFHKPGDGTGTELIGVCGPG-----WA 1818
QY 749 CNVGC 753
DB 1819 ANISC 1823

RESULT 2
VGLX_HSVB STANDARD; PRT; 797 AA.
AC P28968;
DT 01-DEC-1992 (REL. 24, CREATED)
DT 01-DEC-1992 (REL. 24, LAST SEQUENCE UPDATE)
DT 01-DEC-1992 (REL. 24, LAST ANNOTATION UPDATE)
DE GLYCOPROTEIN X PRECURSOR.
GN 71.
OS EQUINE HERPESVIRUS TYPE 1 (STRAIN AB4P) (EHV-1).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; HERPESVIRIDAE; ALPHAHERPESVIRINAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 92295566.
RX TELFORD E.A.R., WATSON M.S., MCBRIDE K., DAIVSON A.J.;
RL VIROLOGY 189:304-316(1992).
DR EMBL; M86664; G330862;
DR FIR; H38802; VGBEX1.
KW GLYCOPROTEIN; TRANSMEMBRANE; SIGNAL.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 797 GLYCOPROTEIN X.
FT DOMAIN 23 465 SER/THR-RICH.
FT TRANSMEM 766 790 POTENTIAL.
FT CARBOHYD 590 590 POTENTIAL.
SQ SEQUENCE 797 AA; 80342 MW; 18DE40E1 CRC32;
```

RA		HAUSER F., HOFFMANN W.:	
RL	J.	BIOLOG. CHEM. 267:24620-24624(1992).	
CC	-!	FUNCTION: COULD BE INVOLVED IN DEFENSE AGAINST MICROBIAL	
CC	-!	INFECTIONS. PROTECTS THE EPITHELIA FROM EXTERNAL ENVIRONMENT.	
CC	-!	SUBCELLULAR LOCATION: SECRETED.	
CC	-!	TISSUE SPECIFICITY: SKIN	
CC	-!	PPM: EXTENSIVELY O-GLYCOSYLATED.	
CC	-!	ALTERNATIVE PRODUCTS: A NUMBER OF DIFFERENT FORMS OF THE PROTEIN	
CC	-!	MAY BE PRODUCED BY ALTERNATIVE SPLICING.	
CC	-!	SIMILARITY: CONTAINS 6 P-TYPE (TREFOL) DOMAINS.	
DR	EMBL;	L02115; G951460; -	
DR	PIR;	A45155; A45155.	
DR	HSSP;	P04002; IATF.	
DR	PROSITE;	PS00025; P_TREFOIL; 6.	
KW	REPEAT;	AMPHIBIAN SKIN; GLYCOPROTEIN; ALTERNATIVE SPLICING.	
FT	NON_TER	1	8 X 8 AA APPROXIMATE TANDEM REPEATS,
FT	DOMAIN	81 144	ALA/THR-RICH.
FT	REPEAT	81 88	1-1.
FT	REPEAT	89 96	1-2.
FT	REPEAT	97 104	1-3.
FT	REPEAT	105 112	1-4.
FT	REPEAT	113 120	1-5.
FT	REPEAT	121 128	1-6.
FT	REPEAT	129 136	1-7.
FT	REPEAT	137 144	1-8.
FT	DOMAIN	161 202	P-TYPE 1.
FT	DOMAIN	218 301	8 X APPROXIMATE TANDEM REPEATS, THR-RICH
FT	REPEAT	218 224	2-1.
FT	REPEAT	225 239	2-2.
FT	REPEAT	240 249	2-3.
FT	REPEAT	250 259	2-4.
FT	REPEAT	260 275	2-5.
FT	REPEAT	276 287	2-6.
FT	REPEAT	288 294	2-7.
FT	REPEAT	295 301	2-8.
FT	DOMAIN	306 347	P-TYPE 2.
FT	DOMAIN	353 394	P-TYPE 3.
FT	DOMAIN	402 522	12 X APPROXIMATE TANDEM REPEATS,
FT			THR-RICH.
FT	REPEAT	402 411	3-1.
FT	REPEAT	412 419	3-2.
FT	REPEAT	420 431	3-3.
FT	REPEAT	432 443	3-4.
FT	REPEAT	444 453	3-5.
FT	REPEAT	454 460	3-6.
FT	REPEAT	461 472	3-7.
FT	REPEAT	473 479	3-8.
FT	REPEAT	480 491	3-9.
FT	REPEAT	492 498	3-10.
FT	REPEAT	499 515	3-11.
FT	REPEAT	516 522	3-12.
FT	DOMAIN	525 566	P-TYPE 4.
FT	DOMAIN	572 613	P-TYPE 5.
FT	DOMAIN	620 661	P-TYPE 6.
FT	DISULFID	162 188	BY SIMILARITY.
FT	DISULFID	172 187	BY SIMILARITY.
FT	DISULFID	182 199	BY SIMILARITY.
FT	DISULFID	307 333	BY SIMILARITY.
FT	DISULFID	317 332	BY SIMILARITY.
FT	DISULFID	327 344	BY SIMILARITY.
FT	DISULFID	354 380	BY SIMILARITY.
FT	DISULFID	364 379	BY SIMILARITY.
FT	DISULFID	374 391	BY SIMILARITY.
FT	DISULFID	526 552	BY SIMILARITY.
FT	DISULFID	536 551	BY SIMILARITY.
FT	DISULFID	546 563	BY SIMILARITY.
FT	DISULFID	573 599	BY SIMILARITY.
FT	DISULFID	583 598	BY SIMILARITY.
FT	DISULFID	593 610	BY SIMILARITY.
FT	DISULFID	621 647	BY SIMILARITY.
FT	DISULFID	631 646	BY SIMILARITY.
FT	DISULFID	641 658	BY SIMILARITY.

RESULT	4
CPN_DROME	

Query Match 10.0%; Score 431; DB 1; Length 865;

Query Match 10.0%; Score 431; DB 1; Length 865;  
Best Local Similarity 25.8%; Pred. No. 1.8e-12;  
Matches 183; Conservative 43; Mismatches 248; Indels 234; Gaps 32;

Qy	91	TLGPPAETTQAPAT---	---TQAPTTTQAPTTTQAPT---	-----TTTQAPTTTT	133
		:    :	:    :	:    :	
Db	5	TIPSPVAPVAPVPSAVAAPVQVVS	PAAPVAPAPAPAIATVPAPPP	TLASQVAPATVI	64
Qy	134	QAPITTOA---	---PTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTT	189	
		:    :	:    :	:    :	
Db	65	PAPAPITAAASVT	PVASVAPPVVAAAPTTPPAASPVST---	---PVAQAIPVASVAPV---	APPV 118
Qy	190	TQAPTTITQ---AATTPAATTP---	---AATTPAATTPAATTPAATTPGAPAPTSAPVMPPICE	244	
		:    :	:    :	:    :	
Db	119	AATPTPVQVQIPVAPVATATPPV	ASASAPTAAVTPVISPVIASPPVVPANTTPVAA	PVA-	178
Qy	245	LLPNGCPADFDIHL	LPHDKYCNLFQCSNGYTFEORCEGLFYFPYVORCDSPANVECD	304	
		:    :	:    :	:    :	
Db	178	AVPAAVFPV-----	-----	---VAPVIAVAPVAPVAVV	202
Qy	305	GETSPAPPVVTEGNEDEDIDIGD	LLDNGCCPANFEIDLWLLPHGNCRCD	RYQCXVHGNLVERRC	364
		:    :	:    :	:    :	
Db	203	AETPAPPVPAE-----	IPVATIPCEVAP-----	LIPEVS-----	232

WRESCHNER D.H., HAREUVENI M., TSARFATY I., SHORODINSKY N.,  
HOREV J., ZARETSKY J., KOTKES P., WEISS M., LATHE R., DION A.,  
KEYDAR I.;  
EUR. J. BIOCHEM. 189:463-473(1990).

[6]  
SEQUENCE FROM N.A.  
TISSUE-BREAST CARCINOMA;  
MEDLINE; 90276414.  
HAREUVENI M., TSARFATY I., ZARETSKY J., KOTKES P., HOREV J.,  
ZRIHAN S., WEISS M., GREEN S., LATHE R., KEYDAR I., WRESCHNER D.H.;  
EUR. J. BIOCHEM. 189:475-486(1990).

[7]  
SEQUENCE FROM N.A.  
MEDLINE; 91033045.  
TSARFATY I., HAREUVENI M., HOREV J., ZARETSKY J., WEISS M.,  
JELTSCH J.M., GARNIER J.M., LATHE R., KEYDAR I., WRESCHNER D.H.;  
GENE 93:313-318(1990).

[8]  
PARTIAL SEQUENCE FROM N.A.  
MEDLINE; 88330762.  
GENDLER S.J., TAYLOR-PAPADIMITRIOU J., DUHIG T., ROTHBARD J.,  
BURCHALL J.;  
J. BIOL. CHEM. 263:12820-12823(1988).

[9]  
SEQUENCE OF 1-46 FROM N.A.  
TISSUE-BREAST CARCINOMA;  
BULUWELA L., LIU Q., LUOMANI Y.A., GOMM J.J., COOMES R.C.;  
SUBMITTED (OCT-1992) TO EMBL/GENBANK/DDBJ DATA BANKS.

-!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CYTOSKELETON.

-!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM  
IS ALSO PRODUCED.

-!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL  
TUMORS, SUCH AS BREAST CANCER.

-!- PTM: HIGHLY GLYCOSYLATED (N- & O-LINKED CARBOHYDRATES & STALIC  
ACID).

-!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
VARIES AMONG DIFFERENT ALLELES. THE BREAST TUMOR MUC1 HAS  
GENERALLY ONLY 5 TO 7 REPEATS.

-!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE  
SPLICING OF THE MUC1 GENE.

EMBL; J05582; G189599; -  
EMBL; M32738; G182124; -  
EMBL; M32739; G182129; -  
EMBL; J05581; G188870; -  
EMBL; M61170; G189784; -  
EMBL; X52229; G37054; ALT\_SEQ.  
EMBL; X52228; G36435; ALT\_SEQ.  
EMBL; M35093; G182253; ALT\_SEQ.  
EMBL; Z17324; G34807; -  
EMBL; Z17325; G34809; -  
EMBL; M21868; G386963; ALT\_SEQ.  
PIR; A35175; A35175.  
PIR; B35175; B35175.  
PIR; S10218; S10218.  
MM; 158340; -

GLYCOPROTEIN; SIGNAL; CYTOSKELETON; ACTIN-BINDING; TRANSMEMBRANE;  
REPEAT; ALTERNATIVE SPLICING; POLYMORPHISM.

SIGNAL	1	23	POTENTIAL.
CHAIN	24	1255	MUCIN 1.
DOMAIN	24	1162	EXTRACELLULAR (POTENTIAL).
TRANSMEM	1163	1186	POTENTIAL.
DOMAIN	1187	1255	CYTOPLASMIC (POTENTIAL).
DOMAIN	81	960	44 X 20 AA TANDEM REPEATS.
CARBOHYD	957	957	POTENTIAL.
CARBOHYD	975	975	POTENTIAL.
CARBOHYD	1029	1029	POTENTIAL.
CARBOHYD	1055	1055	POTENTIAL.
CARBOHYD	1133	1133	POTENTIAL.
VARSPLIC	19	19	T -> ATPAPKAT (IN FORM B).
VARSPLIC	20	22	MISSING (IN FORM C).
VARSPLIC	20	31	MISSING (IN FORM D).
VARSPLIC	126	905	MISSING (IN POLYMORPHIC EPITHELIAL



RESULT	6
AMYH_YEARST	
ID AMYH_YEARST	STANDARD;
AC P08640; P08068;	PRT; 1367 AA.
DT 01-AUG-1988 (REL. 08, CREATED)	
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)	
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)	

[illegible]

```
Db 523 SSAPAPTPSSSTTESSAPV----- 543
QY 365 GAGTHFSELOQCQDHIELVGLTLPGESEEVVDVEDACTGWYCPTPIEWELPNGCPAD 424
Db 543 -----TSSTTESSAPV-----PTPSSSTTE 563
QY 425 FSIDHLLPHESDCQYLOCVHGQTIARPCGNLHFSFATOSCSPVTAAGCOVFECDSDNQ 484
Db 564 SS-----STPVT-----SSTTESSAPVPT-----PSSST 588
QY 485 CTSTAAPTAAP-----TAAPTAAPTAAPSTVVVPATPDATAPV-----PT 530
Db 589 TESSAPVPTPSSSTTESSAPAPTPSSSTTESSAPVPTSSSTTESSAPVPTSSSTTES 648
QY 531 TAIPTPAPTAAPTAAAPESPTTVTPPTAAPTAAPTAVPEIPIT-----VTSAPT 585
Db 649 SSAPVPTPSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTES 707
QY 586 APTAAPT-----AAPTAAPTAVPE-----IPTVTSPTAAPTAAAPNTVTVPT 635
Db 708 PTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTSSSTTES 767
QY 636 AAPT-----AAPAPN-TTVTVPTAAPTAAAP-----PTVAHAPNTAAPTAAPT 685
Db 768 SSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTESSAPVPTPSSSTTES 827
QY 686 -----EDDDIDPPLND-----PINPCVECNVLP 710
Db 828 FSSSTESSVPVPTPSSSTTESSAPVSSSTTESSAPV 865

RESULT 7
SLP1_CLOTM STANDARD; PRT; 1664 AA.
AC Q06852;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
GN OLBP.
OS CLOSTRIDIUM THERMOCELLUM.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NCIB 10682;
RX MEDLINE: 93209931.
RA FUJINO T., BEGUIN P., AUBERT J.-P.;
RL J. BACTERIOL. 175:1891-1899(1993).
CC -1- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -1- SUBCELLULAR LOCATION: CELL WALL.
CC -1- SIMILARITY: CONTAINS 4 S-LAYER HOMOLGY (SLH) DOMAINS.
DR EMBL: X67506; G296881;
DR PROSITE: PS01072; SLH_DOMAIN; 2.
KW CELL WALL; S-LAYER; SIGNAL; REPEAT.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
FT T-P-S-D-E-P.
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
FT DOMAIN 1495 1565 SLH 2.
FT DOMAIN 1566 1625 SLH 3.
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
SQ SEQUENCE: 1664 AA; 178194 MW; BPEC7D6A CRC32;
```

```
Query Match 8.2%; Score 353.5; DB 1; Length 1664;
Best Local Similarity 23.5%; Pred. No. 6.2e-09;
Matches 149; Conservative 70; Mismatches 277; Indels 139; Gaps 23;

QY 93 PGPAETTAQAPATTAQTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQ 152
Db 917 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 976
QY 153 APTTTTQ-----APTPTTQAP-----TTTTQAAPTTTQAAPTTTQAAPTTTQAAPT 202
Db 977 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPS 1036
QY 203 PAATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 262
Db 1037 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD----- 1087
QY 263 DKYCNLFYQCSNGYTFEQRCEGLYFNYPVQRCDSPANVECDGEIAPAPVTEGNEDEDI 322
Db 1087 -----EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1122
QY 323 DIGDLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCGAGTHFSFELQCCDHIEL 382
Db 1123 DEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD-----EP 1150
QY 383 VGCITLPGGSEEVVDVEDACTGWYCPTPIEWELPNGCPADFSIDHLLPHESDCQYLO 442
Db 1151 TPSPDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD----- 1204
QY 443 CVHGQTIARPCGNLHFSAP-----TQSCSPVTAAGCOVFECDSDNQCTSTAAPTAA 497
Db 1204 -----EPTPSD-EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD 1252
QY 498 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 557
Db 1253 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD----- 1311
QY 558 TVPPTAAPTAAPTAVPEIPITVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAVPEIPT 616
Db 1311 ---PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD----- 1355
QY 617 APTTAAPAPNTVTVPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 662
Db 1356 DEPT---PSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSD-----P 662
QY 663 TVAHAPNTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 697
Db 1413 TVPTSTP-----TPTSKPTSTAPTEIPEPTPSD 1442

RESULT 8
VTP3_TTV1V STANDARD; PRT; 474 AA.
AC P19275;
DT 01-NOV-1990 (REL. 16, CREATED)
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)
DE VIRAL PROTEIN TPX.
OS THERMOPROTEUS TENAX VIRUS 1 (STRAIN VT3) (TTV1).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; LIPOTHRIVIRIDAE.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE: 90245666.
RA NEUMANN H., ZILLIG W.;
RL NUCLEIC ACIDS RES. 18:2171-2171(1990).
DR EMBL: X14717; E1192689; -.
DR PIR: S06886; S06886.
DR PIR: S15921; S15921.
KW REPEAT.
FT DOMAIN 278 467 3 THR-PRO REPEATS REGIONS AND TWO NEAR
FT IDENTICAL REPEATS.
FT REPEAT 278 367 THR-PRO(N).
FT REPEAT 368 377 THR-PRO(N).
FT REPEAT 378 436 THR-PRO(N).
```

```
FT REPEAT 437 446
FT REPEAT 447 467 THR-PRO(N).
SQ SEQUENCE 474 AA; 49507 MW; BAC5D831 CRC32;

Query Match 8.18; Score 351; DB 1; Length 474;
Best Local Similarity 32.68; Pred. No. 3.1e-09;
Matches 103; Conservative 20; Mismatches 129; Indels 64; Gaps 13;

Qy 394 EVDVDEACTGWCTEPIEWELPNCGPADFSIDHLLPHESDCGOYLQCVHGOTIARPC 453
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 192 EIDISQFAAF-----TDPLVMYLSINGSTANVKV-----YKQGSNIGTVSGNYSTPY 239

Qy 454 PGN-----LHFSFATQSCSPVTAAGCQVFECDNQCTSTTAAPTAAP 495
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 240 -GNFSMAGYGTVDKHYANFIVLPYEDPQVTVPI-----SSPSPTPTPTPTPT 288

Qy 496 TAAATAAPTAAPTAAPS---TVVPATPPATAAPVPTTAITPTAPTAAPTAAPTAAP 552
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 289 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 348

Qy 553 SPT-TVTVPPTAAPTAAPTTAVPEIPTVTSAPTAAPTAAPTAAPTTAVPEIPT-T 610
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 349 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 402

Qy 611 VTSPPTAAPTAAAPNTTVPPTAAPT-TAAPAPNTTPT-----VPPTAAPTAAPTV 664
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 403 TPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 456

Qy 665 AHAPNTAAPTPTTSA 680
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 457 TPTPTPTPTPTSTSS 472

RESULT 9
ID Y091_NVPVP STANDARD; PRT; 279 AA.
AC O10341;
DT 01-NOV-1997 (REL. 35, CREATED)
DT 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE HYPOTHETICAL 29.3 KD PROTEIN (ORF92).
OS ORGIA PSEUDOTSUGATA MULTICAPSID POLYHEDROSIS VIRUS (OPMNPV).
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; BACULOVIRIDAE; EUBACULOVIRINAE.
RN [1]
RZ SEQUENCE FROM N.A.
RX MEDLINE; 9271300.
RA AURENS C.H.; RUSSELL R.R.; FUNK C.J.; EVANS J.; HARWOOD S.,
RA ROHRMANN G.F.;
RL VIROLOGY 229:381-399(1997).
CC -!- SIMILARITY: TO CORRESPONDING ORF IN ACMNPV.
DR EMBL; U75930; G1911338;
KW HYPOTHETICAL PROTEIN.
SQ SEQUENCE 279 AA; 29289 MW; AAB6954 CRC32;

Query Match 8.18; Score 350.5; DB 1; Length 279;
Best Local Similarity 34.28; Pred. No. 2.2e-09;
Matches 102; Conservative 34; Mismatches 113; Indels 49; Gaps 14;

Qy 484 QCTSTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 543
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 26 ONIEKPLPSPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 82

Qy 544 AAPTAAPESTTVPPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 603
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 83 PSLTSLPSPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 126

Qy 604 VPEIPTTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 661
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 127 ---PTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 181

Qy 662 -PTVAHAPNTAAPTPTTTTSAAPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 720
```

```
Db 182 SPTPPSPPLGDPMPYFPSSVGTQDLQDYIRPLCNGWPRP---RETYAVPW---DCQVRM 236
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Qy 721 VCDGNQVL--VVCSEGLQFNPTTKTCDFAACNVGCVRSNIQMSSEYGVQVFIIPWNL 776
   I : : : : : I : : : : : I : : : : : I : : : : : I : : : : :
Db 237 HCSYFSLPLHEVTCNDDSAISFTAD-----GCV-----PHYOSDCPFYPLNTL 279

RESULT 10
ANP_NOTCO
ID ANP_NOTCO STANDARD; PRT; 822 AA.
AC P24856;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ANTIFREEZE GLYCOPOLYMER POLYPROTEIN AFGP7/AFGP8 PRECURSOR.
GN AFGP8.
OS NOTORHENTIA CORIICEPS NEGLECTA (BLACK ROCKCOD) (YELLOWBELLY ROCKCOD).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC OSTEICHTHYES; ACTINOPTERYGII; PERCIFORMES.
RN [1]
RZ SEQUENCE FROM N.A.
RX MEDLINE; 91067687.
RA HSIAO K.-C.; CHENG C.-H.C.; FERNANDES I.E.; DETRICH H.W. III,
RA DEVRIES A.L.;
RL PROC. NATL. ACAD. SCI. U.S.A. 87:9265-9269(1990).
CC -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE LIVER AND SECRETED INTO
CC THE BLOOD FROM WHICH THEY BECOME DISTRIBUTED TO ALMOST THE ENTIRE
CC EXTRACELLULAR SPACE.
CC -!- PTM: THE DISACCHARIDE GALACTOSE-N-ACETYL GALACTOSAMINE IS ATTACHED
CC TO THREONINES IN AFGP8 AND AFGP7.
CC -!- CONTAINS 44 COPIES OF AFGP8 AND TWO COPIES OF AFGP7.
DR EMBL; M55000; G213392;
DR PIR; A38420; A38420.
DR HSP; P04002; IATF.
KW ANTIFREEZE PROTEIN; GLYCOPOLYMER; POLYPROTEIN; REPEAT;
KW MULTIGENE FAMILY; SIGNAL.
RN [1]
RZ SIGNAL 1 ?
RZ PROPEP 37
RZ CHAIN 38 822
RZ PEPTIDE 38 822 ANTIFREEZE GLYCOPOLYMER POLYPROTEIN.
RZ PEPTIDE 55 68 AFGP8.
RZ PEPTIDE 72 85 AFGP8.
RZ PEPTIDE 89 102 AFGP8.
RZ PEPTIDE 106 119 AFGP8.
RZ PEPTIDE 123 136 AFGP8.
RZ PEPTIDE 140 153 AFGP8.
RZ PEPTIDE 157 170 AFGP8.
RZ PEPTIDE 174 187 AFGP8.
RZ PEPTIDE 191 204 AFGP8.
RZ PEPTIDE 208 221 AFGP8.
RZ PEPTIDE 225 238 AFGP8.
RZ PEPTIDE 242 255 AFGP8.
RZ PEPTIDE 259 272 AFGP8.
RZ PEPTIDE 276 289 AFGP8.
RZ PEPTIDE 293 306 AFGP8.
RZ PEPTIDE 310 323 AFGP8.
RZ PEPTIDE 327 340 AFGP8.
RZ PEPTIDE 344 357 AFGP8.
RZ PEPTIDE 361 374 AFGP8.
RZ PEPTIDE 378 391 AFGP8.
RZ PEPTIDE 395 408 AFGP8.
RZ PEPTIDE 412 425 AFGP8.
RZ PEPTIDE 429 442 AFGP8.
RZ PEPTIDE 446 459 AFGP8.
RZ PEPTIDE 463 476 AFGP8.
RZ PEPTIDE 480 493 AFGP8.
RZ PEPTIDE 497 510 AFGP8.
RZ PEPTIDE 514 527 AFGP8.
RZ PEPTIDE 531 544 AFGP8.
RZ PEPTIDE 548 561 AFGP8.
RZ PEPTIDE 565 578 AFGP8.
```

[illegible]



[illegible]

```

RL SUBMITTED (SEP-1995) TO EMBL/GENBANK/DDBJ DATA BANKS.
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR
CC      (POTENTIAL).
CC -1- PTM: EXTENSIVELY O-GLYCOSYLATED (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE SRP1/TIPI FAMILY OF STRESS-INDUCED
CC      PROTEINS.
CC EMBL; Z49651; G1015903; -.
DR PROSITE; PS00724; SRP1.TIPI; 1.
KW HYDROTHERMAL PROTEIN; GLYCOPROTEIN; SIGNAL; GPI-ANCHOR.
FT SIGNAL          1       24     POTENTIAL.
FT CHAIN           25    1146   HYPOTHETICAL PROTEIN YR151C.
FT PROPEP         1147    1161   REMOVED IN MATURE FORM (POTENTIAL).
FT LIPID           1146    1146   GPI-ANCHOR (POTENTIAL).
SQ SEQUENCE        1161 AA; 118358 MW; 55125063 CRC32;

Query Match              7.1%; Score 309; DB 1; Length 1161;
Best Local Similarity 19.5%; Pred. No. 3.8e-07;
Matches 120; Conservative 70; Mismatches 224; Indels 200; Ga

Qy 73 TEKFSAQTCHVAALAGCTLPGEPAETTQAPATTQAPTPTTQAPTPTTQAPTPTT
Db | | | | : | | | | | | | | | | | | | | | | | | | | | | | |
128 TTRASSSTTPPTITSTTTSTTPTTSSTT--PTTSTTSTPTTSTTSTPTTSTT
Qy 133 TQAPTPTTQAPTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQA
Db : | | | | | | | | | | | | | | | | | | | | | | | | | | |
186 STTPTTS---TTSTPTTSTTSTPTTSTTPTTSTTPTTSTTPTTSTTPTTSTTSQT
Qy 193 PTTITQATTPAATTAAATTPAATTAAATTPGVPAPTASPVWPPICELLPNCPA
Db | | | | | | | | | | | | | | | | | | | | | | | | | | |
243 SKTSTPTTSTSTPTTSTPTTSTTSTTSTTSTTSTTSTTSTTSTT-----
Qy 253 DFDHLLPHDKYNLFYOCNSNGYTFEQRCEGLYFNPYVORCDSPANVECDGEISAPP

```

```
Db 287 -----ISTAPT 292
Qy 313 VTEGNEDEIDIGDLLONGCPANFEIDWLLPHGNKCDKYYQCVHGNLVERRCGAGTHFSF 372
Db 293 TST-----TSTTF 300
Qy 373 ELQOCDHIELVGCPLPGSESEVDVEDACTGWCPTPEIEMWELPNCPCPADFSIDHLLP 432
Db 301 STSSASASSVISTT-----ATTSTTFASLITTPATSTASTDHTS 339
Qy 433 HESDCGOYLQCVHQTARPCPNLHSPATQSCSPVTAGCQVFCDSNOCTSTAAPT 492
Db 340 SVSTNAF-----TTSATTTTSDTYI-SSSSPSQVTSSEPT 376
Qy 493 AAPTAAPTAAPTAAPTAAPSTVPPATPPATAAPVPPPTAL-PTPAPTAAPTAAPTAAP 551
Db 377 TVSEVTSVEPTR-----SSQVTSSEPTVSEF--TSSVEPTRSSQVTSSEPTTV-- 427
Qy 552 ESPTTVTPPTAAPTAAPTTAAPVPEIPITVTSAPTAAPTAAPTAAPTAAPTTAAPVPEIPTV 611
Db 427 -SEPTSSVEPTRSS-----QVTSSAEPITVSEFTSSVEPTRSSQV-----TSS 468
Qy 612 TSPPTAAPTAAAPNTVIVPPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 671
Db 469 AEPTVSEFTSSVEPTRSSQVTSSEPTVSEFTSSVEPTRSSQVTSSEPTVSEFTSS 528
Qy 672 AAPVTTSAPATTP 685
Db 529 VEPIRSSQVITTEP 542

RESULT 15
CH12_COCIM STANDARD; PRT; 860 AA.
ID P54197;
AC P54197; 1996 (REL. 34, CREATED)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DT 01-OCT-1996 (REL. 34, LAST SEQUENCE UPDATE)
DE ENDOCHITINASE 2 PRECURSOR (EC 3.2.1.14).
GN CTS2.
OS COCCIDIOIDES IMMITIS.
OC EUKARYOTA; FUNGI; DEUTEROMYCOTINA (IMPERFECT FUNGI).
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C735;
RX MEDLINE: 96144270.
RA PISHKO E.J., KIRKLAND T.N., COLE G.T.;
RL GENE 167:173-177(1995).
CC -1- FUNCTION: MAY BE ASSOCIATED WITH ENDOSPORULATION.
CC -1- CATALYTIC ACTIVITY: HYDROLYSIS OF THE 1,4-BETA-LINKAGES OF
CC N-ACETYL-D-GLUCOSAMINE POLYMERS OF CHITIN.
CC -1- SIMILARITY: BELONGS TO CHITINASE CLASS II (FAMILY 18 OF GLYCOSYL
CC HYDROLASES).
DR EMBL: L41662; GI200192; -.
DR PROSITE; PS01095; CHITINASE_18; 1.
KW HYDROLASE; GLYCOSIDASE; CHITIN DEGRADATION; CHITIN-BINDING; SIGNAL;
KW GLYCOPROTEIN.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 860 ENDOCHITINASE 2.
FT CARBOHYD 90 90 POTENTIAL.
SQ SEQUENCE 860 AA; 91395 MW; 00EC2B0 CRC32;
```

Query Match 7.1%; Score 308; DB 1; Length 860;  
Best Local Similarity 22.3%; Pred. No. 3.3e-07;  
Matches 143; Conservative 71; Mismatches 220; Indels 206; Gaps 22;

```
Qy 93 PGPAETTTQAPATTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 148
Db 342 PDPP--TSTVTSASSTSTQSSQSTMETKLSASTPSSPSVSPSTM-QTTSTGS 398
Qy 149 TTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 206
```

```
Db 399 TSIEVTTRSQEPSTTISTRASSTEFVTTRSQEPSTTISTRASSTEFVTTRSQEPST 458
Qy 207 TPAATTAAATTPAAT---TPAATTGCVPAPTSAPVWPPICELLPNGCPADFDHLLIPH 262
Db 459 T--ISTWSASTESTSQSPSTTISTKSAPT----- 489
Qy 263 DKYCNLFYQCSNGYTFEQRCEGLYFNYPYQRCDSPANVECDGEISAPPVWTEGNEDEI 322
Db 489 -----GTVTTRSQDLPSSTISTRSPETET 512
Qy 323 DIGDLLONGCPANFEIDWLLPHGNKCDKYYQCVHGNLVERRCGAGTHFSFELQOCDHIEL 382
Db 513 ETATTKSQGPSI-----TLSTRSSA-----ET 536
Qy 383 VGCTLPGGESEEDVDVEDACTGWCPTPEIEMWELPNCPCPADFSIDHLLPHESDCGOYLQ 442
Db 537 VSTRSQSHSSSTTSTKS-----APTE-----TGTSEHST----- 567
Qy 443 CVHGQTIARPCPNLHSPATQSCSPVTAGCQVFCDSNOCTSTAAPTAAPTAAPTAA 502
Db 567 -----SMPVSTRSASTETVIT---RSQNSDSQSMVSTRSPS---TESITR 607
Qy 503 PTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 562
Db 608 SQGPSSETFTSKSPVVDITISTELPSQTPITITGTP-SDPVSAPT-----TVPPN 657
Qy 563 AAPTAAPTAAVPEIPTVTS-APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTT 621
Db 658 PTTLAPSSSTEDRTTITITISVTVCTGTTTITITITITITITITITITITITITIT 717
Qy 622 AAPAP-----NTTVTPPTAAPTAAAPNT-TVTVPPPTAAPTAAPTAAHAPNTTAA 675
Db 718 GAPAPPDGWTITVCPQCAPT-----PTTVTLTVPTRSA--FLPARTETRPVTVVPV 770
Qy 676 -----TTTSAPATTPEDDDIDPPLPND 697
Db 771 PENPIKNVKPSESQDFVTVTAAAPATVTKTLEYNNPVDSD 810
```

Search completed: July 18, 1999, 01:41:31  
Job time: 2314 sec



Result No.	Score	Query %			DB	ID	Description
		Match	Length	Match			
1	636	14.7	3020	2	A43932	mucin 2 precursor,	
2	440	10.2	873	2	A47283	calphatin - fruit	
3	433	10.0	797	1	VG8EX1	glycoprotein X pre	
4	432	10.0	662	2	A43155	mucin FIM-C.1 - Af	
5	431	10.0	865	2	A47282	calcium-binding pr	
6	397	9.2	1335	2	A35687	mucin precursor, p	
7	397	9.2	1255	2	B35175	episialin B - huma	
8	397	9.2	1295	2	A35686	polymorphic epithe	
9	397	9.2	1264	2	A35175	episialin A - huma	
10	393.5	9.1	1367	2	A8478	glucan 1.4-alpha-g	
11	392	9.1	1188	2	S45915	extensin-like prot	
12	381	8.8	294	2	A37232	mucin, tracheal (A	
13	361.5	8.4	1118	2	A8292	mucin, tracheobron	
14	354	8.2	622	2	A38420	antifreeze glycopr	
15	351	8.1	474	2	S15921	protein TPX-VT3 -	
16	345	8.0	232	2	A60095	larval glue protei	
17	340.5	7.9	353	2	S36438	EPP7 protein - hyd	
18	340	7.9	279	2	S53363	mucin 5AC (clone J	
19	339.5	7.8	901	2	A49227	sialidase - Actino	
20	331.5	7.7	400	2	A28172	spasmolysin precu	
21	330.5	7.6	670	2	F36791	hypothetical prote	
22	328	7.6	379	2	S50125	larval 'glue' protei	
23	324	7.5	216	2	I47141	gastric mucin (clo	
24	322	7.4	528	2	S15920	mucin - rhesus mac	
25	321.5	7.4	350	2	S24556	hydroxyproline-ric	
26	314.5	7.3	303	2	S28456	hydroxyproline-ric	
27	314	7.3	328	2	QJ0985	hydroxyproline-ric	
28	309.5	7.2	327	2	S20074	promastigote surfa	
29	309	7.1	907	1	QJBE21	membrane antigen 9	
30	309	7.1	1161	2	S57180	probable membrane	
31	308.5	7.1	451	2	S71754	cellular hepatitis	
32	308	7.1	860	2	JC4566	chitinase (EC 3.2.	
33	307.5	7.1	307	1	GF5F3	salivary glue prot	
34	306.5	7.1	1630	2	A53577	ascites sialoglyco	
35	306	7.1	267	2	S09314	cell wall glycopro	
36	305.5	7.1	700	2	A54641	interspersed repea	
37	303.5	7.0	377	2	A48018	mucin 7 precursor,	
38	303	7.0	447	2	S49321	mucin - rat (fragm	
39	301.5	7.0	886	2	S29605	glycoprotein 350/2	

A:Molecule type: DNA  
A:Residues: 1343-1352, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:gl188863; PID:gl188864  
A>Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:55749, NCBI:P:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lampert, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evl  
A:Reference number: A33532; MUID:89197956  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:gl188873; PID:gl188874  
A:Experimental source: intestine  
R:Jany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.  
J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.  
A:Reference number: A61257  
A:Accession: A61257  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: T, 1925-1948, 'TTS', 1952-1954 <JAN>  
A:Experimental source: bronchus  
R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCool, D.; Jones, C.; Forstn  
Biochem. Biophys. Res. Commun. 183, 821-828, 1992





[illegible]

Query Match	10.0%	Score 431;	DB 2;	Length 865;
Best Local Similarity	25.8%	Pred. No. 5.9e-15;		
Matches 183;	Conservative 43;	Mismatches 248;	Indels 234;	Gaps 32;

  

QY	91	TLPGPPAETTQAPAT---	TOAPT	TTQAPT	TTTTQAPT	-----	TTTQAPT	TTTT	133
							:		
Db	5	TIPSPVAPVAAAPT	PSA	VAAAPQVWSPAA	VAPAPAPAPIA	VT	PVAPP	PTLASVQA	PTVI 64
QY	134	QAPTTTQA	----	PTTQAPT	TTQAPT	TTTTQAPT	TTTTQAPT	TTTTQAPT	TTTQAPT
Db	65	PAPAPIAAASVTP	VAS	VAPVPPVAAAPT	PTPAAS	PSVST---	PVAAQ	ITPVAS	APV---APPV 118

Qy	190	TQAPTITIQ-AANTPAATP-----AATTAAATTAATTPGVPAPTSAPVWPPICE	244
Dy	119	AATPTPVQIPVAAPVIATPPVAASAPTAAATPVISPVIASPPVPANTTVPAAAPVA-	178
Qy	245	LLPNGCPADFDIHLIPHDKYCNLFQCSNGYTFEQRCPEGLFYFNYYQRCDSPANVECD	304
Dy	178	AVPAAPVV-----VAPVLAPVAPAPVAPVV	202
Qy	305	GEISPAPPVTGNEDEDEDIGDLLDNGCSPANFEIDMLPHGNRCDKRYOCVHGNLVERRC	364
Dy	203	AETPAPPVPAE-----IPVATIPECVAP-----LIPEVS-----	232
Qy	365	GAGTHSFELQQCDHIELVGCTTLPGGESSEVDVEDACTGWICTPTPEIWEPLNCCPAD	424
Dy	232	-----VVATKLAAAEPVVVAPPA-----TE-----TPVVAPAA	260
Qy	425	FSIDLHPHSDCGQYLQCQVHGOTIARPCPNLHFSPATQSCBSPVTAGCQVFECDSNDQ	484
Dy	261	AS-----PHYS-----VAPAVETAAPVPSASTPEPPVAA-----	290
Qy	485	CTSTAATAAPTAAPTAA-----PFAATAAPSTVVPP-----	518
Dy	290	ATLTAP-ETPALAPVVAESQVAANTVATPPPPAP--EPETIAPPVWAETPEVASVAVA	346
Qy	518	-ATPPATAAPVPTTAIPTAPTAAPTAAPTAAESPITVT-----VPRTAAPTAAPTT	571
Dy	347	ETTPPV-----VPPVAESIPAPVVVATTPVATUAVTDVITASVELPPVIAPSPPV-S	401
Qy	572	AVPEIPITVSAPTAAPTA-----PTAAPTAAPTA-----VPEIPTTTVTSF-	615
Dy	402	AVAETPVDL-APPVLPPVAEEPVAVVAEETPETPAPASAPVTIAALDIPEVAPVIAAPS	460
Qy	615	--PTAAPTAAAP--APNTIVTVPTTAAPTAAAPANTTVTPPTAA---PTAAPTVAH	666
Dy	461	DAPAEAPSAAPVISTVPTTASVPETTPAAPVPTPIDVSVLSAAIETPVAPPVEV---	519
Qy	667	APNTAAPTMTTSAPATTPEDDDIDPPLPNDPNCPCVECNLPMAHA	714
Dy	519	--TTEVAVADV-APPEAAADLIIEPVEPPAPIPDLLLEQTTSVPAYEA	562
RESULT	6		
A35887		mucin precursor, pancreatic - human	
C:		Species: Homo sapiens (man)	
C:		Date: 23-Oct-1990 #sequence_revision 23-Oct-1990 #text_change 03-Feb-1999	
C:		Accession: A35887	
R:		Lan, M.S.; Batra, S.K.; Qi, W.N.; Metzgar, R.S.; Hollingsworth, M.A.	
J:		Biol. Chem. 265, 15294-15299, 1990	
A:		Title: Cloning and sequencing of a human pancreatic tumor mucin cDNA.	
A:		Reference number: A35887; MUID:90368716	
A:		Accession: A35887	
A:		Status: preliminary; not compared with conceptual translation	
A:		Molecule type: mRNA	
A:		Residues: 1-1335 <LAN>	
A:		Cross-references: GB:J05582	
C:		Superfamily: polymorphic epithelial mucin	

```

Query Match          9.2%; Score 397; DB 2; Length 1335;
Best Local Similarity 26.1%; Pred. No. 4.1e-13;
Matches 199; Conservative 42; Mismatches 288; Indels 234; Gaps

QY 10 ALGIVA--ARPEVSDAEKNPALHEPHDPCPPASQHWLLPHEYDCTFYCEYGLKFIAPR 67
      |||      |      |      |      |      |      |      |      |
Db 107 ALGSTTPPAHDVTSAPDNKPA---PGSTAPPA-----HGVTSAPTD 144

QY 68 DCAPGTETKEFSQCTCVHAAAGCTLCGPPAETTTQAPATTTTQAPTTT-----TQ 120
      |||      |||      |      |      |      |      |      |
Db 145 RPAGS-----TAPPAHGVTSAPTDRPAGSTAPPAHGVTSAPTDRPAGHGVTS 200

QY 121 APTTTTQAPTTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 179

```



QY 68 DCAPGTEKFSQATCVHAALAGCTLPGPPAETTQAPATTOAATTQAPTTT-----TQ 120  
Db 145 RPAGS-----TAPPAHGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTS 200  
QY 121 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTQAPTTTQAPT 179  
Db 201 APDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTS 260  
QY 180 APTTQAPTQAPT-TTQAAAT---PAATPAA---TTPAATTAAATTAAATTGVA 232  
Db 261 APDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTS 317  
QY 233 PTSAP-----VWPPICELLPNCPADFDIHLIPHDKYCNLFYQCSNGYTF--EQR 282  
Db 318 VTSAPDTRPAGSTAPP-----AHGVTSAPDTRP 346  
QY 283 PEGLYFNYPVQRCDSANVECDGEISAPPVTEGNEDEDIDIGDLLDNCGPANFEIDWLL 342  
Db 347 APGSTAPPAGVTSAPDTRPAG--STAPP----- 375  
QY 343 PHGNRCDKYQCVHGNLVERRCGAGTHFSFELQCCDHIELVGTLPGGSEEEVDVDEDAC 402  
Db 375 -----AHG-----VTSAPDTRPAGSTAPPAGVTSAPDTRP 407  
QY 403 TGWYCP-----TEPIEWELPNCPADFSIDHLLPHESDCGOYLOCVHGQTIA---R 454  
Db 408 PGSTAPPAGVTSAPDTRPAG-----GSTAPPAGVTSAPDTRPAP 448  
QY 455 GNLHFSPATQSCSPVACQVFECDSDNOCTSTAAPTAAPTAAPTAAPTAAPTSTV 514  
Db 449 GST-APPAGVTSAP-----DTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 497  
QY 515 V--PPATPPATAAPVPP---TTAIPPTAAPTAAPTAAPTAAPTAAPTAAPTAA 569  
Db 498 VTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 557  
QY 570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTAAPTAAPTAAPT 616  
Db 558 VTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 617  
QY 616 -TAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 664  
Db 618 VTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGST 671  
QY 664 -----VAHAPNTTAAPTVTT-----TSAPATTPEDDDDIDPP 693  
Db 672 APPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPP 714

## RESULT 9

A35175  
episialin A - human  
C:Species: Homo sapiens (man)  
C>Date: 10-Aug-1990 #sequence\_revision 10-Aug-1990 #text\_change 23-Feb-1997  
C:Accession: A35175  
R:Lightenberg, M.J.L.; Vos, H.L.; Gennissen, A.M.C.; Hilken, J.  
J. Biol. Chem. 265, 5573-5578, 1990  
A:Title: Episialin, a carcinoma-associated mucin, is generated by a polymorphic gene end  
A:Reference number: A35175; MUID:90202794  
A:Accession: A35175  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1264 <LIG>  
A:Cross-references: GB:M32738; GB:M34088; GB:J05288  
C:Superfamily: polymorphic epithelial mucin  
C:Keywords: cytoskeleton; transmembrane protein

Query Match 9.2%; Score 397; DB 2; Length 1264;  
Best Local Similarity 26.1%; Pred No. 3.9e-13;  
Matches 199; Conservative 42; Mismatches 288; Indels 234; Gaps 34;

QY 10 ALGLVA--ARPEVSDAEKNPALHEPHDCPPABQHWLLPHEPDCTKFYCYEXGLKFIAPR 67

Db 116 ALGSTTPPAHGVTSAPDNKPA---PGSTAPPA-----HGVTAPD 153  
QY 68 DCAPGTEKFSQATCVHAALAGCTLPGPPAETTQAPATTOAATTQAPTTT-----TQ 120  
Db 154 RPAGS-----TAPPAHGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTS 209  
QY 121 APTTTTQAPTTTQAPTTTQAPTTTQAPT-TTTQAPTTTQAPTTTQAPT 179  
Db 210 APDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTS 269  
QY 180 APTTQAPTQAPT-TTQAAAT---PAATPAA---TTPAATTAAATTAAATTGVA 232  
Db 270 APDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 326  
QY 233 PTSAP-----VWPPICELLPNCPADFDIHLIPHDKYCNLFYQCSNGYTF--EQR 282  
Db 327 VTSAPDTRPAGSTAPP-----AHGVTSAPDTRP 355  
QY 283 PEGLYFNYPVQRCDSANVECDGEISAPPVTEGNEDEDIDIGDLLDNCGPANFEIDWLL 342  
Db 356 APGSTAPPAGVTSAPDTRPAG--STAPP----- 384  
QY 343 PHGNRCDKYQCVHGNLVERRCGAGTHFSFELQCCDHIELVGTLPGGSEEEVDVDEDAC 402  
Db 384 -----AHG-----VTSAPDTRPAGSTAPPAGVTSAPDTRP 416  
QY 403 TGWYCP-----TEPIEWELPNCPADFSIDHLLPHESDCGOYLOCVHGQTIA---R 454  
Db 417 PGSTAPPAGVTSAPDTRPAG-----GSTAPPAGVTSAPDTRPAP 457  
QY 455 GNLHFSPATQSCSPVACQVFECDSDNOCTSTAAPTAAPTAAPTAAPTAAPTAA 514  
Db 458 GST-APPAGVTSAP-----DTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 506  
QY 515 V--PPATPPATAAPVPP---TTAIPPTAAPTAAPTAAPTAAPTAAPTAAPTAA 569  
Db 507 VTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 566  
QY 570 TTAVPEI-PITVTSAP-----TAAPTAAPTAAPTAAPTAAPTAAPTAAPT 616  
Db 567 VTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPPAG 626  
QY 616 -TAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 664  
Db 627 VTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGS--TAPPAHGVTSAPDTRPAGST 680  
QY 664 -----VAHAPNTTAAPTVTT-----TSAPATTPEDDDDIDPP 693  
Db 681 APPAGVTSAPDTRPAGSTAPPAGVTSAPDTRPAGSTAPP 723

## RESULT 10

S48478  
glucan 1,4-alpha-glucosidase (EC 3.2.1.3) - yeast (Saccharomyces cerevisiae)  
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein YIR01  
C:Species: Saccharomyces cerevisiae  
C>Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 05-Dec-1997  
C:Accession: S48478; A26877; B26877; S27281; JC6123  
R:Rowley, K.  
submitted to the EMBL Data Library, October 1994  
A:Reference number: S48478  
A:Accession: S48478  
A:Molecule type: DNA  
A:Residues: 1-1367 <ROW>  
A:Cross-references: GB:Z47047; EMBL:Z38061; MIPS:YIR019c; NID:g603997; PID:g763364  
J:Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
A:Reference number: A91831; MUID:87194600  
A:Accession: A26877  
A:Molecule type: DNA  
A:Residues: 1-242 <YAM>

A:Cross-references: EMBL:M16164; NID:g172522; PID:g172525  
A:Accession: B26877  
A:Molecule type: DNA  
A:Residues: 762-1331 <YAX>  
A:Cross-references: EMBL:M16165; NID:g172523; PID:g172526  
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Clarios, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Saccharomyces cerevisiae  
A:Reference number: S27281  
A:Accession: S27281  
A:Molecule type: DNA  
A:Residues: 1-31 <PAR>  
A:Cross-references: EMBL:X13957; NID:g4551; PID:g4552  
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A:Title: Mucl, a mucin-like protein that is regulated by Msl0, is critical for pseudohyphal growth of *Saccharomyces cerevisiae*  
A:Reference number: JC6123  
A:Accession: JC6123  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1367 <LAM>  
A:Cross-references: GB:U30626; NID:g1304386; PID:g1304387  
C:Genetics:  
A:Gene: SGD:MUC1; STA2; MAL5; DEX2  
A:Cross-references: MIPS:YIR019C; SGD:S0001458  
A:Map position: 9R  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosidase  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F:5-21/Domain: transmembrane #status predicted <TM1>  
F:1350-1366/Domain: transmembrane #status predicted <TM2>

Query Match 9.1%; Score 393.5; DB 2; Length 1367;  
Best Local Similarity 19.8%; Pred. No. 6.3e-13;  
Matches 150; Conservative 120; Mismatches 229; Indels 259; Gaps 28;

QY 32 PHDCCPAQOHLLPHEYDCTKFEYCEYGLKFTAPDRCAFGTEKFSATCVHAALAGCT 91  
DB 288 PHHDTTP-----CTK-----KKTTSKTCIKKTTTPVTPSSSTESSAP 328  
QY 92 LGGPPAETQ---APATTOAPTTOAPTITTOAPT-----TTOAPTITTOAPTITQAP 142  
DB 329 VPTPSSSTESSAPVTSSTESSAPVTPSSSTESSAPVTSSTESSAPVTSST 388  
QY 143 TTTQAPTITTOAPTITTOAPTITTOAPTITTOAPTITTOAPTITTOAPTIT 196  
DB 389 ESSAPVPTSSSTESSAPVTSSTESSAPVTSSTESSAPVTSSTESSAPVTS 448  
QY 196 -TQOATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 245  
DB 449 STTESSAPVPTSSSTESSAPVTSSTESSAPVTSSTESSAPVTSSTESSAPV 501  
QY 246 LPNGCCADFDIHLIIPHDKYCNLFYOCNGYFEQRCPEGLFENYVQRCDSANVECDG 305  
DB 501 -----TSSTESSAP-----TSSTESSAP-----VPTPSSSTES 522  
QY 306 EISAPAP-PVTEGNEDEIDIGDLLDNGCPANFEIDLPHGNRCDKYQCVHGNLVERRC 364  
DB 523 SSAPATPSSSTESSAPV-----TSSTESSAP-----TSSTESSAP----- 543  
QY 365 GAGTHFSFELQOCDHIELVGCITLPGGESEVVDDEACTGWICTPEIWEPLNGCPAD 424  
DB 543 -----TSSTESSAPV-----TSSTESSAPV-----PTPSSSTTE 563  
QY 425 FSTDHLLPHESDCGQVLCVHGQTIARPCGNLHFSFATQSCESPVTAGCQVFECDSDNQ 484  
DB 564 SS-----STPV-----STTESSAPVPT-----PSSST 588  
QY 485 CTSTAATTAAP-----TAATTAATTAATTAATTAATTAATTAATTAATTAAT 530  
DB 589 TESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTES 648  
QY 531 TAIPTPAATTAATTAAPESPTTVTPPTAATTAATTAATTAATTAATTAATTAAT 585

DB 649 SSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPV 707  
QY 586 AAPTAAPT---AAPTAAPTAAVPE-----IPTVTGSPPTAAPTAAAPNTTIVPT 635  
DB 708 PTPSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPT 767  
QY 636 AAPT---AAPT---AAPT---AAPT---AAPT---AAPT---AAPT---AAPT--- 685  
DB 768 SSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSS 827  
QY 686 -----EDDDIDPPLND-----PINPCVECNLP 710  
DB 828 FSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSSSTESSAPVPTSS 865  
RESULT 11  
S49915  
extensin-like protein - maize  
C:Species: Zea mays (maize)  
C:Date: 05-Mar-1995 #sequence\_revision 12-May-1995 #text\_change 09-Sep-1997  
C:Accession: S49915  
R:Rubinstein, A.L.; Broadwater, A.H.; Lowrey, K.; Bedinger, P.A.  
submitted to the EMBL Data Library, June 1994  
A:Description: Pex genes: pollen-specific genes with extensin-like domains.  
A:Reference number: S49915  
A:Accession: S49915  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1188 <RUB>  
A:Cross-references: EMBL:Z34465; NID:g600117; PID:g600118

Query Match 9.1%; Score 392; DB 2; Length 1188;  
Best Local Similarity 22.5%; Pred. No. 6.6e-13;  
Matches 170; Conservative 87; Mismatches 324; Indels 176; Gaps 28;

QY 2 IKTLFLTALGLVAARPEYSDAKNPALHPEHPDC-----PPAEQHLLPHEYDCTKFY 56  
DB 520 VKTSPAPIGSPPPPPVSVVPPVVKSPPPPPAPVGVSPPPPEKSPPPP----- 570  
QY 57 CEYGLKFTAPDRCAFGTEKFSATCVHAALAGCTLPGPAETTOAPTQAP----- 110  
DB 570 -----APVASPPPPVVKSPPPPTLVASPPPPVVKSPPPPPAPVASPPPPV 621  
QY 110 -TTQAPTITTOAPTITTOAPTITTOAPTITTOAPTITTOAPTITTOAPTIT 168  
DB 622 SPPPPAPVASSPPPMKSPPPPTVSSPPPEKSPPPPPPAKSTPPPEYPTTSVKSSP 681  
QY 169 TTTQAPTITTOAPTITTOAPTITTOAPTITTOAPTITTOAPTITTOAPTIT 215  
DB 682 PPEKSLPPTLIPSPPPQEKPTTPTSPKPPSPPEKSPPEKSPPEKSPPPPPV 741  
QY 215 -----ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 262  
DB 742 SSPPPTPVSSPPALAPVSSPPVVKSPPPPPAPLSSP--PPAPQV--KSSPPVQVSSPP 797  
QY 263 DKYCNLFYOCNGYFEQRC-PEGLFENYVQRCDSANVECDGELSPAPPVTEGNEDE 321  
DB 798 PKSSPPPLAPVSSPPQVKTSPPPAPLSSPPAPLSSPPAPLSSPPVSSPPV 852  
QY 322 IDIGDLLDNGCPANFEIDLPHGNRCDKYQCVHGNLVERRCGAGTHFSFELQOCDHIE 381  
DB 852 -----PAPVSSPPPLTP-----KPAASPPAHVS-----SPPE 876  
QY 382 LVGCTLPGGESEVVDDEACTGWICTPEIWEPLNGCPADFSIDLHLLPHESDCGQV 441  
DB 877 VVKPSTPAPTIVISP-----PSEPKSSPP-----PTPVSLPPPVIKSS----- 916  
QY 442 QCVHGQTIARPCGNLHFSFAT-QSCESPVTAGCQVFECDSDNCTSTAAPTAATTA 500  
DB 916 -----PPAMVSSPPMTPKSPPPVWV-----SSPPPTVKSSPPPPVSSPP 957



```

Qy      629 TTVVPPTAAPTAAAPNTTIVPPTAAPTAAPT 663
       ||||| :||| :||| :||| :||| :|||
Db      263 TTTVTPTPTSTGRTGTPTTSITTTTTVTP - --PT 294

RESULT 13
A48292
mucin, tracheobronchial - dog
C:Species: Canis lupus familiaris (dog)
C:Date: 27-Jun-1994 #sequence_revision 27-Jun-1994 #text_change 10-Sep-1997
C:Accession: A48292
R:Verma, M.; Davidson, E.A.
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993
A:Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA con
A:Reference number: A48292
A:Accession: A48292
A>Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1118 <VER>
A:Cross-references: GB:X69164; NID:g402557; PID:g402558
A:Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as
C:Comment: It is uncertain whether Met-1 or Met-52 is the initiator.
C:Keywords: glycoprotein

Query Match          8.4%; Score 361.5; DB 2; Length 1118;
Best Local Similarity 23.3%; Pred. No. 2.1e-11;
Matches 207; Conservative 49; Mismatches 295; Indels 339; Gaps 37;

Qy      66 PRDCAPGETERKFSQTCVHAALAGCTLP-----GPPAETTQAPATTQAPTTTQAP 115
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |
Db      203 PPQCLTPTPTPLSPSIIRPOPEALTLPCLSAVSETTPTISITTTTPTPTPTGTQT 262
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      116 TTTTQAFTTTTQAFTTT--TQAFTT-----TQ 140
       ||||| :||| :||| :||| :||| :|||
Db      263 TTTPTTTTPTPTPTGTGTPTTLVLTITTTTMTPTPTSQKLRGSDTPVLVIALVTPT 322
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      141 APTTQAFTTTQAFTTTTQAFTT---TQAFTTTQAFTTTQAFT----- 183
       ||||| :||| :||| :||| :||| :|||
Db      323 TPTGTPTPTPTTTTIVTPTPTPTGTGTPTTIRITTTTMTPTPTTSTKVLVTPYTT 382
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      183 -----TQAFTTTQAFTTTTQAFTTQAFTTQAFTPA-----ATTCAA 220
       ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      383 TTGVPTTHRHRTQTTHTTTYTVAIATPTGTRPTTVLHTTTTQTTHRTTREPQ 442
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      221 TTPAATTGPVPAPTSAPWPPICELLNGCPADFIHLIPHDKYCNLFQCNSGYTFEQ 280
       ||||| :||| :||| :||| :||| :|||
Db      443 TTPSPTTTVD-PQTQHPSPPLGLDPNEN-THRDTDPN----HDTTTT--TVTPTPTPR 494
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      281 RCPEGLFNPVQRCDSPANVECDGEISPAPPYTEGNE-----EDI 322
       || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      495 TRPOPRIYTITTTMTPTP-----TPPTHFDHHDLPPPPPVRRHHRTPRHCHEL 545
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      323 DIGDLL--DNGCPANFEIDMLPHGNRCOKYQCVH-----GNLVERRCGAG 367
       ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      546 TSGSPLHYPTPTTN-----LLEKRNSRLINTPTTWIISSAMATPPSLVHSRCRG 599
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      368 THSFEELOQ-----CDHIEL-----VGCTLPGG-----ESEEVDVEDAC 402
       ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      600 -HSSKAORKSRASGGLMMLGVESGVGVSTHAGGALVRSCRFRTHSRAE----- 651
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      403 TGWCPTPEPIWEPLNGCCPADFSIDILLPHES-----DCGOYLQCVHQGTIARPCPG 455
       ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      651 -----GLGPEAVYSMLGVESGVGVGRHCGRGLRSAGGG--AEDDPG 691
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      456 NL-----HFSPATQSCSPV-----TAGQVFECDSNOCTSAAPTA 493
       ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||   ||
Db      692 RVPRVGALTDSTQHTSKAIKCKAGVNEELTTYIACASALQSETCAADMIATATIHTT 751
       |   |   |   |   |   |   |   |   |   |   |   |   |   |   |   |

Qy      494 APT-----AAPTAAPTAAPASTVV-----PPATPATAAPVPT 530
       || :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      752 SRSRKPKTISELITYACASTAASGVGWVOLDMIAITGGLRD RGOMAGAMPAGSRGPSFG 811

```

531 TAIPPTAPTAAPTAAPTAAPESPPTTVTPVPTAAPTAAPTTAVPEIPITVTSAPTAAPTA 590  
 Db 812 ELEFPNGLOAPTTPIS-----TTTIVTPTPTG-----TQTPTTTPSP 852  
 Qy 591 APTAAPTAAPTT--AVPEIPTVTSPTTAAPT----- 622  
 Db 853 PPGDPTKGKVVYKIIGSLRTTFNPLPAWLGVITICKLMQSNRKTKTAETINVKTLGSLM 912  
 Qy 622 --AAPAPNTTVTP-----PTA-----APT-----TAA 642  
 Db 913 MDVTPINSCRVTPTPPSGKPCADSRMLNPTAETGTVGAGCCCVLTPTPTHVRIRCTTA 972  
 Qy 643 PAPNTTVTPPTAAPTAAPTTVAHAPNTTAAPVTTTTSAPATTPEDDDIDP 692  
 Db 973 PTP-TPISITTTTVPQTQHP--AHRP--TTSPITTTTT-VTQPDTHTDP 1016

RESULT 14  
 A38420  
 antifreeze glycoprotein precursor - black rockcod  
 C:Species: Notothenia coriiceps neglecta (black rockcod)  
 C:Date: 28-Jun-1991 #sequence\_revision 30-Jan-1993 #text\_change 09-Sep-1997  
 C:Accession: A38420  
 R:Hisao, K.C.; Cheng, C.H.C.; Fernandes, I.E.; Detrich, H.W.; DeVries, A.L.  
 Proc. Natl. Acad. Sci. U.S.A. 87, 9285-9269, 1990  
 A:Title: An antifreeze glycopeptide gene from the antarctic cod Notothenia coriiceps neg  
 A:Reference number: A38420; MUID:91067687  
 A:Accession: A38420  
 A>Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-822 <HSI>  
 A:Cross-references: GB:M55000; NID:g213391; PID:g213392  
 A:Note: the authors translated the codon CCT for residue 489 as Ala  
 C:Keywords: glycoprotein

Query Match	8.28; Score 354; DB 2; Length 822;
Best Local Similarity	25.28; Pred. No. 3.8e-11;
Matches	185; Conservative 21; Mismatches 319; Indels 210; Gaps 25;
QY 65	APRCARGTEPKFSAQTCVHAA-----LACTLPGPP-----AETQAPATTOQAPT 111
DB 146	AATPATPAFHAATAATPATAATPALIFAATAATAATPATPAFHAATAATPATAATPAL 205
QY 112	TOAPTTTTQAA-PTTITQAPTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPT 170
DB 206	IFAATAATAATPAALNFAATAATPATAATPALIFAATAATAATPATAALNFAATAATA 265
QY 171	TTOAPTTTQAPTTTQAPTTTQAPTTTQAA-----ATTPA-----ATTPA-ATTP- 214
DB 266	ATPATACNFNAATAATPATAATPALIFAATAATAATPATACNFNAATAATPATAATPALI 325
QY 214	-AATTPAATTA-----ATTPGPVAPTSAVPWPPICELLPGNCPADFDHLLPHDKYC 266
DB 326	FAATAATAATPATAALNFAATAATPATAATP-----ATTPA-----ATTPA-ATTP- 357
QY 267	NLFYQCSNGYTFEQRCEGLYFNYPVORCDSPANVECDGEISPAAPPVTEGNEDEDIDGD 326
DB 357	-----ALIFAATAATAATPATAALNFAATAATPATAAT-PAALIFAAT 397
QY 327	LLDNGCPANFEIDWLLPHGNRCDKYOCVHGNLVERCGAGT-----HFSFELQOCDHIEL 382
DB 398	AATAATPATAAL-----NFAATAATAATPATPAFNF-----AA 430
QY 383	VGCTLPGGESEVDVEDACTGWYCPTPIEWEPLNGCCPADFSIDHLLPHESDCGOYLQ 442
DB 431	TAATPATAATPALIFAATAATA-----ATPATAALNFAATAATPATAATPALIF 479
QY 443	CVHGQTIARPCPGNLHFSFATQSCSPVTAGCQVFECDSNNQCTSTAAPTAAPATAA 502
DB 480	AATAATAATPATAALHFAATAATAATPATAAL-----NFAATAATPATAATPALIFA 531

Qy	503	PTAAPTAAPST-----VVPPTPPATA-----APVPPTTAITP--APTAAPTAAP	546
		:	
Db	532	ATAATAATPATAAFNAFAATAATAATPATAALNFAATAATAATPATAALIFAATAATAATP	591
		:	
Qy	547	TTAAPESTTVVPPTAAPTA-----APTAVPEIP-----	578
		:	
Db	552	ATAALNFAATAATPATAATPALIFAATAATAATPATPAFHFAATAATPATAATPALIFAA	651
		:	
Qy	578	--IIVTSAPTAAPTAAPTAAPTTAVPEIPTTVSPPTAA--PTTAAPAPNTVTVVPP	634
		:	
Db	652	TAATAATPATAALNFAATAATAATPATAALNFAATAATAATPATAALIFAATAATAATPA	711
		:	
Qy	635	TAA-----PTTAA-----PAPNTVTVVPTA-----APTAPPTVAHAP----	669
		:	
Db	712	TAAALNFAATAATPATAATPALIFAATAATAATPATAAFNAFAATAATAATPATAATPALIF	771
		:	
Qy	669	NTTAAPVTTTISAPAT	683
		:	
Db	772	AATAATAATPATPAT	786
		:	
RESULT 15			
S15921			
C:Species: Thermoproteus phase 1 (strain strain KRAL 10/12)			
A:Variety: Thermoproteus phase 1, TTV1			
C:Variety: strain KRAL 10/12			
C:Date: 21-Nov-1993 #sequence_revision 26-May-1995 #text_change 09-Sep-1997			
C:Accession: S15921; S06686			
R:Neumann, H.: Zillig, W.			
Nucleic Acids Res. 18, 2171, 1990			
A:Title: Nucleotide sequence of the viral protein TPX of the TTV1 variant VT3.			
A:Reference number: S15921; MUID:90245666			

A: Molecule type: DNA  
A: Residues: 1-474 <NEW>  
A: Cross-references: EMBL: X14717; NID: g62191; PID: e20711; PID: g584349  
A: Experimental source: strain KRAL 10/12  
A: Note: the authors translated the initiation codon GTG for residue 1 as Val  
R: Neumann, H.  
submitted to the EMBL Data Library, March 1989  
A: Reference number: S06686  
A: Accession: S06686  
A: Molecule type: DNA  
A: Residues: 207-474 <NEW>  
A: Cross-references: EMBL: X14717  
A: Experimental source: strain KRAL 10/12  
C: Genetics:  
A: Start codon: GTG

[illegible]

Qy 665 AHAPNTTAAAPVTTTSA 680  
| | | :|||  
Db 457 TPTPTPTPTSTTSS 472

Search completed: July 18, 1999, 00:45:05  
Job time: 261 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 03:21:16 ; Search time 42.46 Seconds  
(without alignments)  
901.494 Million cell updates/sec

Title: US-09-294-663-3

Perfect score: 4328

Sequence: 1 MIKTLFLTALGLVARPEV.....VFIPWNKLDIEDIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 396714 seqs, 48575446 residues

Database : Pending\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/paa/PCURS9\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US07\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US080\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US081\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US082\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US083\_COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US084A\_COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US084B\_COMB.pep.\*
- 9: /cgn2\_6/ptodata/1/paa/US085\_COMB.pep.\*
- 10: /cgn2\_6/ptodata/1/paa/US086\_COMB.pep.\*
- 11: /cgn2\_6/ptodata/1/paa/US087\_COMB.pep.\*
- 12: /cgn2\_6/ptodata/1/paa/US088\_COMB.pep.\*
- 13: /cgn2\_6/ptodata/1/paa/US089\_COMB.pep.\*
- 14: /cgn2\_6/ptodata/1/paa/US090\_COMB.pep.\*
- 15: /cgn2\_6/ptodata/1/paa/US091\_COMB.pep.\*
- 16: /cgn2\_6/ptodata/1/paa/US092\_COMB.pep.\*
- 17: /cgn2\_6/ptodata/1/paa/US093\_COMB.pep.\*
- 18: /cgn2\_6/ptodata/1/paa/US094\_COMB.pep.\*
- 19: /cgn2\_6/ptodata/1/paa/US095\_COMB.pep.\*
- 20: /cgn2\_6/ptodata/1/paa/US096\_COMB.pep.\*
- 21: /cgn2\_6/ptodata/1/paa/US097\_COMB.pep.\*
- 22: /cgn2\_6/ptodata/1/paa/US098\_COMB.pep.\*
- 23: /cgn2\_6/ptodata/1/paa/US099\_COMB.pep.\*
- 24: /cgn2\_6/ptodata/1/paa/US100\_COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4328	100.0	788	16	US-09-294-663-3
2	4293	99.2	786	15	US-09-103-429A-3
3	4266.5	98.6	807	16	US-09-294-663-4
4	4231.5	97.8	805	15	US-09-103-429A-4
5	521.5	12.0	1205	2	US-07-643-502C-28
6	520.5	12.0	1404	2	US-07-643-502C-13
7	520.5	12.0	1404	2	US-07-757-022B-2
8	520.5	12.0	941	2	US-07-757-022B-14
9	520.5	12.0	1361	2	US-07-757-022B-40
10	520.5	12.0	1311	2	US-07-757-022B-42
11	520.5	12.0	1270	2	US-07-757-022B-44
12	520.5	12.0	1320	2	US-07-757-022B-46
13	520.5	12.0	1354	2	US-07-757-022B-48
14	520.5	12.0	1314	2	US-07-757-022B-50
15	520.5	12.0	1363	2	US-07-757-022B-52
16	520.5	12.0	1049	2	US-07-757-022B-58
17	520.5	12.0	1320	2	US-07-757-022B-60
18	520.5	12.0	1404	2	US-07-757-022B-62
19	520.5	12.0	1038	2	US-07-757-022B-74
20	520.5	12.0	1022	2	US-07-757-022B-84

21	520.5	12.0	1140	2	US-07-757-022B-104
22	520.5	12.0	1313	2	US-07-757-022B-142
23	455.5	10.5	1837	13	US-08-928-361B-5
24	455.5	10.5	1837	21	US-08-928-361B-5
25	445.5	10.3	1721	1	PCT-US97-14104-5
26	445.5	10.3	1721	11	US-08-700-651A-5
27	445.5	10.3	1721	11	US-08-700-651-5
28	445.5	10.3	1721	13	US-08-928-361B-6
29	445.5	10.3	1721	21	US-08-928-361B-6
30	393.5	9.1	1367	18	US-60-087-236-55
31	373.5	8.6	557	16	US-09-248-796-26892
32	373.5	8.6	557	18	US-60-096-409-26892
33	333.5	7.7	827	16	US-09-248-796-17307
34	333.5	7.7	827	18	US-60-096-409-17307
35	322.5	7.5	3880	3	US-08-028-021-1
36	317	7.3	297	9	US-08-580-545B-6
37	317	7.3	297	16	US-09-262-653A-6
38	313	7.2	907	11	US-08-783-774-2
39	309.5	7.2	216	13	US-08-928-361B-8
40	309.5	7.2	216	21	US-08-928-361B-8
41	301	7.0	216	21	US-08-928-361B-27
42	298.5	6.9	288	16	US-09-216-393-341
43	298.5	6.9	288	16	US-09-216-393-344
44	296	6.8	249	11	US-08-700-651-15
45	296	6.8	249	13	US-08-928-361-20

ALIGNMENTS

RESULT 1  
US-09-294-663-3  
Sequence 3, Application US/09294663  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
TITLE OF INVENTION: CDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
STREET: 118 North Tioga Street  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/294,663  
FILING DATE: 19-APR-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/103,429  
FILING DATE: 24-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 3:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 788 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO

;  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic membrane  
US-09-294-663-3

Query Match 100.0%; Score 4328; DB 16; Length 788;  
Best Local Similarity 100.0%; Pred. No. 1e-271;  
Matches 788; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MIKTLFLTALGLVAARPVSDAENKPNALHEPHDPCCPAEQHLLPHEVDCTKFYCEYG 60  
DB 1 MIKTLFLTALGLVAARPVSDAENKPNALHEPHDPCCPAEQHLLPHEVDCTKFYCEYG 60  
QY 61 LKFIAPRDCAPGTEKFSQAQTCVHAALAGCTLPGPAETTQAPATTQAPTTTQAPTTTQ 120  
DB 61 LKFIAPRDCAPGTEKFSQAQTCVHAALAGCTLPGPAETTQAPATTQAPTTTQAPTTTQ 120  
QY 121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
DB 121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
QY 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 240  
DB 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 240  
QY 241 PICELLPNGCPADFDIHLIPHDKYNLFYQCSNGYTFEQRCPGELYFNYPYVQRCDS 300  
DB 241 PICELLPNGCPADFDIHLIPHDKYNLFYQCSNGYTFEQRCPGELYFNYPYVQRCDS 300  
QY 301 VECDEISAPPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
DB 301 VECDEISAPPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
QY 361 ERRCGAGTHFSELOQCDHIELVGTCLPGGESEEDVDDEDACTGWYCTPEIWEPLNG 420  
DB 361 ERRCGAGTHFSELOQCDHIELVGTCLPGGESEEDVDDEDACTGWYCTPEIWEPLNG 420  
QY 421 CPADFSIDLPHESDCGOYLQCVHGOTIARPCGNLHFSPTQSCSPVTAGCQVFECD 480  
DB 421 CPADFSIDLPHESDCGOYLQCVHGOTIARPCGNLHFSPTQSCSPVTAGCQVFECD 480  
QY 481 SDNQCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540  
DB 481 SDNQCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540  
QY 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
DB 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
QY 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 660  
DB 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 660  
QY 661 PPTVAHAPNTTAAPTTTTAPATTDEDDIDPPLNDPINCVEECNVLWPAHADCKY 720  
DB 661 PPTVAHAPNTTAAPTTTTAPATTDEDDIDPPLNDPINCVEECNVLWPAHADCKY 720  
QY 721 VCDGNQVLVWVSEGLQFNPTTKTODFACNVCGRSNIOMSESIEGVQVFIIPWKLDE 780  
DB 721 VCDGNQVLVWVSEGLQFNPTTKTODFACNVCGRSNIOMSESIEGVQVFIIPWKLDE 780  
QY 781 ROALNFEL 788  
DB 781 ROALNFEL 788

RESULT 2  
US-09-103-429A-3  
; Sequence 3. Application US/09103429A  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R

;  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; ORGANISM: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 786 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic membrane  
US-09-103-429A-3

Query Match 99.2%; Score 4293; DB 15; Length 786;  
Best Local Similarity 99.7%; Pred. No. 1.9e-269;  
Matches 786; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

QY 1 MIKTLFLTALGLVAARPEVSDAENKPNALHEPHDPCCPAEQHLLPHEVDCTKFYCEYG 60  
DB 1 MIKTLFLTALGLVAARPEVSDAENKPNALHEPHDPCCPAEQHLLPHEVDCTKFYCEYG 60  
QY 61 LKFIAPRDCAPGTEKFSQAQTCVHAALAGCTLPGPAETTQAPATTQAPTTTQAPTTTQ 120  
DB 61 LKFIAPRDCAPGTEKFSQAQTCVHAALAGCTLPGPAETTQAPATTQAPTTTQAPTTTQ 120  
QY 121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
DB 121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
QY 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 240  
DB 179 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 238  
QY 241 PICELLPNGCPADFDIHLIPHDKYNLFYQCSNGYTFEQRCPGELYFNYPYVQRCDS 300  
DB 239 PICELLPNGCPADFDIHLIPHDKYNLFYQCSNGYTFEQRCPGELYFNYPYVQRCDS 298  
QY 301 VECDEISAPPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
DB 299 VECDEISAPPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 358  
QY 361 ERRCGAGTHFSELOQCDHIELVGTCLPGGESEEDVDDEDACTGWYCTPEIWEPLNG 420  
DB 359 ERRCGAGTHFSELOQCDHIELVGTCLPGGESEEDVDDEDACTGWYCTPEIWEPLNG 418

QY 421 CPADFIDHLLPHESDCGOYLQCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480  
Db 419 CPADFIDHLLPHESDCGOYLQCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 478  
QY 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540  
Db 479 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 538  
QY 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
Db 539 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 598  
QY 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 560  
Db 599 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 558  
QY 661 PPTVAHAPNTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 720  
Db 659 PPTVAHAPNTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 718  
QY 721 VCDGNNOVLVVCSEGLQFNPTKTCDACNVGCVRSNIQMSSEYEGVQVFIIPWKLDE 780  
Db 719 VCDGNNOVLVVCSEGLQFNPTKTCDACNVGCVRSNIQMSSEYEGVQVFIIPWKLDE 778  
QY 781 ROALNFEL 788  
Db 779 ROALNFEL 786

RESULT 3

US-09-294-663-4  
; Sequence 4, Application US/09294663  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: cDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioga Street  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/294,663  
; FILING DATE: 19-APR-1999  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/103,429  
; FILING DATE: 24-JUN-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,590  
; REFERENCE/DOCKET NUMBER: BTI-39-CIP  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 4:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 807 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO

; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic membrane  
US-09-294-663-4

Query Match 98.6%; Score 4266.5; DB 16; Length 807;  
Best Local Similarity 97.1%; Pred. NO. 9.8e-268;  
Matches 784; Conservative 0; Mismatches 4; Indels 19; Gaps 1;

QY 1 MIKTLLEFALGLVAARPEVSDAEKNPALHEPHDPDPRAEQHWWLLPHEXDCFKFYCEVG 60  
Db 1 MIKTLLEFALGLVAARPEVSDAEKNPALHEPHDPDPRAEQHWWLLPHEXDCFKFYCEVG 60  
QY 61 LKFIAPRCAPGTEKFSQAQTCVHAALAGCTLPGPAAETTQAPATTQAPTTTQAPTTTQ 120  
Db 61 LKFIAPRCAPGTEKFSQAQTCVHAALAGCTLPGPAAETTQAPATTQAPTTTQAPTTTQ 120  
QY 121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
Db 121 APTTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
QY 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 240  
Db 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 240  
QY 241 PICELLPNGCPADFDIHLILPHDKYCNLFYQCSNGYTFEORCEGLYFNPNYVORCDSPAN 300  
Db 241 PICELLPNGCPADFDIHLILPHDKYCNLFYQCSNGYTFEORCEGLYFNPNYVORCDSPAN 300  
QY 301 VECDEGEISPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
Db 301 VECDEGEISPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
QY 361 ERRCAGTHFSFELQCCDHIELVGCITLPGGESEEVVDVEDACTGWYCTEPIEWEPLNG 420  
Db 361 ERRCAGTHFSFELQCCDHIELVGCITLPGGESEEVVDVEDACTGWYCTEPIEWEPLNG 420  
QY 421 CPADFSIDHLLPHESDCGOYLQCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480  
Db 421 CPADFSIDHLLPHESDCGOYLQCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480  
QY 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540  
Db 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540  
QY 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
Db 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
QY 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 641  
Db 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 660  
QY 661 PPTVAHAPNTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 701  
Db 661 PPTVAHAPNTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 720  
QY 702 CVEECNVLPWAHADCKYWCNGNNOVLVVCSEGLQFNPTKTCDACNVGCVRSNIQMS 761  
Db 702 CVEECNVLPWAHADCKYWCNGNNOVLVVCSEGLQFNPTKTCDACNVGCVRSNIQMS 761  
QY 762 ESYEGVQVFIIPWKLDEDIROALNFEL 788  
Db 762 ESYEGVQVFIIPWKLDEDIROALNFEL 807

RESULT 4

US-09-103-429A-4  
; Sequence 4, Application US/09103429A  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R

APPLICANT: Wang, Ping  
TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
CDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
STREET: 118 North Tlaga  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 805 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-4

Query Match 97.8%; Score 4231.5; DB 15; Length 805;  
Best Local Similarity 96.9%; Pred. No. 1.8e-265;  
Matches 782; Conservative 0; Mismatches 4; Indels 21; Gaps 3;

QY 1 MIKTLFLTALGLVAARPEVSDAEKNPALHEPHDPDPPAEQHWLLPHEYDCTKFFYCEYG 60  
DB 1 MIKTLFLTALGLVAARPEVSDAEKNPALHEPHDPDPPAEQXLLPHEYDCTKFFYCEYG 60  
QY 61 LKFIAPRDCAPGTEFKFSQTCVHAALAGCTLPGPAETTTQAPATTQAPTTTQATTQ 120  
DB 61 LKFIAPRDCAPGTEFKFSQTCVHAALAGCTLPGPAETTTQAPATTQAPTTTQATTQ 120  
QY 121 APITTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180  
DB 121 APITTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 178  
QY 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 240  
DB 179 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 238  
QY 241 PICELLPNGCPADFIDHLLPHDKYCNLFYQCSNGYTFEQRCEGLYFNPYVQRCDS 300  
DB 239 PICELLPNGCPADFIDHLLPHDKYCNLFYQCSNGYTFEQRCEGLYFNPYVQRCDS 298  
QY 301 VECDSGELSPAPVTEGNEDEIDIGDLLDNGCPANFEIDWLLPHGNRCRDKYQCVHGN 360  
DB 299 VECDSGELSPAPVTEGNEDEIDIGDLLDNGCPANFEIDWLLPHGNRCRDKYQCVH 358  
QY 361 ERRCGAGTHSFELQOCDHLELVCGLTLPGGSESEVDVDEACTGWCPTPIEWELP 420  
DB 359 ERRCGAGTHSFELQOCDHLELVCGLTLPGGSESEVDVDEACTGWCPTPIEWELP 418

QY 421 CPADFSDHLLPHESDCGOYLQCVHGQTIARPCPNLHESPATQSCSPVTAGCQVFEC 480  
DB 419 CPADFSDHLLPHESDCGOYLQCVHGQTIARPCPNLHESPATQSCSPVTAGCQVFEC 478  
QY 481 SDNQTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540  
DB 479 SDNQTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 538  
QY 541 APTAAPTTAAPESTTTVTPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
DB 539 APTAAPTTAAPESTTTVTPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 598  
QY 601 TTAVPEIPTTTSPTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 641  
DB 599 TTAVPEIPTTTSPTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 658  
QY 642 APAPTNTTVPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 701  
DB 659 APAPTNTTVPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 718  
QY 702 CVEECNVLPWAHADCDKYWVCDGNNOVLVVCSEGLQFNPTTKTCDFACNVGCVRSNI 761  
DB 719 CVEECNVLPWAHADCDKYWVCDGNNOVLVVCSEGLQFNPTTKTCDFACNVGCVRSNI 778  
QY 762 ESYEGVQVFIWNKLDLDIRQALNEEL 788  
DB 779 ESYEGVQVFIWNKLDLDIRQALNEEL 805  
RESULT 5  
US-07-643-502C-28  
; Sequence 28, Application US/07643502C  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: A Megakaryocytopoietic Factor  
; NUMBER OF SEQUENCES: 28  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/643,502C  
; FILING DATE: 19910118  
; CLASSIFICATION: 424  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cseri, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5166D  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 28:



SEQUENCE CHARACTERISTICS:  
LENGTH: 1205 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: unknown  
MOLECULE TYPE: peptide  
US-07-643-502C-28

Query Match 12.0%; Score 521.5; DB 2; Length 1205;  
Best Local Similarity 29.5%; Pred. No. 5.3e-26;  
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;

```
QY 93 PGPPAE-TTQAPATT-----QAPTTTQAPTTT-----QAPTTTQ-APTITTOAP 136
Db 132 PTPKAETTTKGPALTTPKEPTTTTPKEPASTTPKEPTTTIKSAPTTTPKEPATTTKSAP 191
QY 137 TTTO--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ-- 186
Db 192 TTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 251
QY 186 -APTTTQAPTTTQATTPAATTTPAATTTPAATTTPAATTTPAATTTPAATTTPAATTTPA 243
Db 252 PAPTTPKEPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 309
QY 244 ELLPNCGPADFDIHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNYPYQVRCDSANVEC 303
Db 309 -----PAPTIT-----PAPTIT-----PAPTIT-----PAPTIT-----PAPTIT 322
QY 304 DGEISPAVPVTEGNEDEDIDGDLNCGPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
Db 322 -----EPAPTT-----EPAPTT-----EPAPTT-----EPAPTT-----EPAPTT 329
QY 364 CGAGTHFSFELQOCDHIELVGTLPGESEEVVDVEDA-CTGWYCTPEIWEPLPNCGP 422
Db 329 -----KSAPTTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 361
QY 423 ADFSIDHLLPHESDCGQYLCQVHGQTIARPCGNLHFSFATQSCSPVTAGCQVFECDSD 482
Db 362 APTT-----PKEP-----PKEP-----PKEP-----PKEP-----PKEP----- 382
QY 483 NOCTSTAAPTAAPTAAPTAAPTAA-APSTVVPPPTAPATAAPVPPTT---AIPIT-- 536
Db 382 ---PKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 438
QY 536 --PAPTAAPTAAPTAAAPTSTTVTP--PAPTAAPTAAPTAAAPTSTTVTP--PAPTAAPT 589
Db 439 EXPAPTPEELAPTPEPTPTTPEEAPTTPKAAAPNI--PKEPATTTPKEPATTTPKE 496
QY 590 AAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 634
Db 497 PAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKE 556
QY 635 TAAPT--AAPAPNTTVVPPTAAPTAAPTVAH-----APNTTAAAPTVAH----- 679
Db 557 GTAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTP 616
QY 680 ---APATTPEDDDIDP---PLPNDPINP 701
Db 617 DRPAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTPKETAPTTP 643
```

RESULT 6  
US-07-643-502C-13  
Sequence 13, Application US/07643502C  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: A Megakaryocytopenic Factor  
NUMBER OF SEQUENCES: 28  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.

STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/643,502C  
FILING DATE: 19910118  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseir, Juann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 51660  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 13:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1404 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-643-502C-13

Query Match 12.0%; Score 520.5; DB 2; Length 1404;  
Best Local Similarity 29.5%; Pred. No. 7.1e-26;  
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;

```
QY 93 PGPPAE-TTQAPATT-----QAPTTTQAPTTT-----QAPTTTQ-APTITTOAP 136
Db 331 PTPKAETTTKGPALTTPKEPTTTTPKEPASTTPKEPTTTIKSAPTTTPKEPATTTKSAP 390
QY 137 TTTO--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ-- 186
Db 391 TTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 450
QY 186 -APTTTQAPTTTQATTPAATTTPAATTTPAATTTPAATTTPAATTTPAATTTPAATTTP 243
Db 451 PAPTTPKEPTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKE 508
QY 244 ELLPNCGPADFDIHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNYPYQVRCDSANVEC 303
Db 508 -----PAPTIT-----PAPTIT-----PAPTIT-----PAPTIT-----PAPTIT 521
QY 304 DGEISPAVPVTEGNEDEDIDGDLNCGPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363
Db 521 -----EPAPTT-----EPAPTT-----EPAPTT-----EPAPTT-----EPAPTT 528
QY 364 CGAGTHFSFELQOCDHIELVGTLPGESEEVVDVEDA-CTGWYCTPEIWEPLPNCGP 422
Db 528 -----KSAPTTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATTTPKEPATT 560
QY 423 ADFSIDHLLPHESDCGQYLCQVHGQTIARPCGNLHFSFATQSCSPVTAGCQVFECDSD 482
Db 561 APTT-----PKEP-----PKEP-----PKEP-----PKEP-----PKEP----- 581
QY 483 NOCTSTAAPTAAPTAAPTAAPTAA-APSTVVPPPTAPATAAPVPPTT---AIPIT-- 536
```

Db 581 ---PKPAPPTTKEPAPPTTKKDPAPAPKEPAPPTTKETAPPTPKKLPTTTPKLPAPTP 637  
QY 536 --PAPTAAPTAAPTAAAPESPTTVTP-PTAAPTAAPTTAVPEIPITVT---SAPTAAPT 589  
Db 638 EKPAPTPEELAPTTPEPTTPEEPAPPTPKAAAPNT--PKPAPPTPKPAPPTPKPE 695  
QY 590 AAPTAAPTAAPTTAVPEIPTTVTSPTTAAPTAAAP---APNTT-----VTVPP 634  
Db 696 PAPPTTKETAPPTPKGTAPTTLKKEPAPPTPKKPAKELAPPTTKETSTSDKPAPTPK 755  
QY 635 TAAPT--AAPAPNTTVTPPTAAPTAAAPTVAH-----APNTTAAPTTTS 679  
Db 756 GTAPTTKEPAPPTPKPAPPTPKGTAPTTLKKEPAPPTPKKPAKELAPPTTKGPTSTTS 815  
QY 680 ---APATTPEDDDIDP--PLPNDPINP 701  
Db 816 DKPAPTTPKETAPTTPKPAPPTPKKP 842

RESULT 7  
US-07-757-022B-2  
; Sequence 2, Application US/07/757022B  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/757,022B  
; FILING DATE: 19910910  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/643,502  
; FILING DATE: 18-JAN-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/546,114  
; FILING DATE: 29-JUN-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/457,196  
; FILING DATE: 29-DEC-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 07/390,901  
; FILING DATE: 08-AUG-1989  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Cserr, Luann  
; REGISTRATION NUMBER: 31,822  
; REFERENCE/DOCKET NUMBER: GI 5190  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (617)876-1170  
; TELEFAX: (617)876-5851  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1404 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-07-757-022B-2

Query Match 12.0%; Score 520.5; DB 2; Length 1404;  
Best Local Similarity 29.5%; Pred. No. 7.1e-26;  
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;

QY 93 PGPPAE-TTQAPATI---QAPTITQAPITTTI-----QAPTITITQ-APTITITQAP 136  
Db 331 PTPKAEITTTKGALTPKPEPTTTTPKEPASTTPKPEPTTTIKSAPTTPKPAPPTTKSAP 390  
QY 137 TTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ--APTITQ-- 186  
Db 391 TTPKPEAPTTTKEPAPPTPKPAPPTTKSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 450  
QY 186 -APTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQ 243  
Db 451 PAPTTTPKEPTTTPKEPAPPTTKPAPPTTKPAPPTTKPAPPTTKPAPPTTKPAPPTTKPAP 508  
QY 244 ELLPNGCPADFDIHLPHDKYCNLFYQCSNGYTEQRCPEGLYENFYQRCDSPANVPC 303  
Db 508 -----PTTTPKEPSTTPK- 521  
QY 304 DGEISPAPVTEGNEDEDIDGDLONGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERR 363  
Db 521 -----EPAPTTT----- 528  
QY 364 CGAGTHFSELOQCDHIELVGLTPGGESEYDVDEDA-CTGMYCPTETIEWEPLPNGCP 422  
Db 528 -----KSAPTTTKEPAPTTTTSAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAP 560  
QY 423 ADFSIDLPHSHESDCGYLQCVHGQTIARPCGNLHFSATQSCSPVTAGCQVPECDS 482  
Db 561 APTT-----PKP-----APTTPKAPPTT----- 581  
QY 483 NQCTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 536  
Db 581 ---PKPAPPTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTT 637  
QY 536 --PAPTAAPTAAPTAAPTTAVPEIPTTVTP-PTAAPTAAPTTAVPEIPTTVT---SAPTAAPT 589  
Db 638 EKPAPTPEELAPTTPEPTTTPPEEPAPPTPKAAAPNT--PKPAPPTPKPAPPTPKPE 695  
QY 590 AAPTAAPTAAPTTAVPEIPTTVTSPTTAAPTAAAP---APNTT-----VTVPP 634  
Db 696 PAPPTTKETAPTTPKGTAPTTLKKEPAPPTPKKPAKELAPPTTKETSTSDKPAPTPK 755  
QY 635 TAAPT--AAPAPNTTVTPPTAAPTAAAPTVAH-----APNTTAAPTTTS 679  
Db 756 GTAPTTKEPAPPTPKPAPPTPKGTAPTTLKKEPAPPTPKKPAKELAPPTTKGPTSTTS 815  
QY 680 ---APATTPEDDDIDP--PLPNDPINP 701  
Db 816 DKPAPTTPKETAPTTPKPAPPTPKKP 842

RESULT 8  
US-07-757-022B-14  
; Sequence 14, Application US/07/757022B  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible









```
QY 244 ELLPNCGPADFDIHLHPDKYCNLFYQCSNGYTFEQRCEGLYFNYPYQRCDSANVEC 303
Db 458 -----PPTTKPSPTTPK- 471
QY 304 DGEISAPPVTEGNEDEDIDGLDNGCPANFEIDWLLPHGNRCCKYYQCVHGNLVERR 363
Db 471 -----EPAPTTT----- 478
QY 364 CGAGTHFSFELQOCDHIELVGCTLPGESEEVDDVDEDA-CTGWYCTPEIWEPLPNCGP 422
Db 478 -----KSAPTTKKEPAPTTTTPKAAAPNT--PKEPAPTTKKEPAPTTTKEP 510
QY 423 ADFSIDHLLPHESDCGYLQCVHGQTIARPCGNLHFSFATQSCSPVTAGCQVFECDSD 482
Db 511 APPT-----PKEP-----APTTPKAPPT----- 531
QY 483 NOCTSTAAPTAAPTAAPTAAPTAA--APSTVVPPATPPATAAPVPPPT---AIPPT-- 536
Db 531 ---PREPAPTTTKEPAPTTTTPKAPAPKAPAPTTTPKKTATTPPKLTTTPEKLAPTT 587
QY 536 ---PAPTAAPTAAPTAAPTTPVTP-PTAAPTAAPTTAVPEIPTVT---SAPTAAPT 589
Db 588 EKPAPTPEELAPTTPEEPTPTTPEEPAPTTTPKAAAPNT--PKEPAPTTKKEPAPTTTKEP 645
QY 590 AAPTAAPTAAPTAVPEIPTTTSPTTAAPTAA--APNTT-----VTVPP 634
Db 646 PAPTTTKEPAPTTTTPKGTATTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKPAPTTPK 705
QY 635 TAAPT--AAPAPNTTVVPPPTAAPTAAPTVAH-----APNTTAAVPTTTS 679
Db 706 GTAPTTKEPAPTTTKEPAPTTTPKGTATTLKEPAPTTTPKPAKELAPTTTKEPTSTTS 765
QY 680 ---APATTPEDDDIDP--PLNDPIMP 701
Db 766 DKPAPTTTKEPAPTTTPKAPTTTPKKP 792
```

RESULT 14

```
US-07-757-022B-50
; Sequence 50, Application US/07757022B
; GENERAL INFORMATION:
; APPLICANT: Gesner, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
```

```
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseir, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 50:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1314 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-757-022B-50

Query Match 12.0%; Score 520.5; DB 2; Length 1314;
Best Local Similarity 29.5%; Pred. No. 6.7e-26;
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;

QY 93 GPPPAE-TQAPATT---QAPTTTQAPTTT-----QAPTTTQ-APTTTQAP 136
Db 241 PTPKAETTTKGALTPKPEPTTTPKEPASTTPKPEPTTTPKSAAPTTPKEPAPTTTKSAP 300
QY 137 TTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ-- 186
Db 301 TTPKEPAPTTTKEPAPTTTPKPEPAPTTTTPKSAAPTTPKSAAPTTPKSAAPTTPKE 360
QY 186 -APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ--APTQ-- 243
Db 361 PAPTTTKEPAPTTTPKPEPAPTTTPKPEPAPTTTPKPAKAPTTTPKPA--PKEA-- 418
QY 244 ELLPNCGPADFDIHLHPDKYCNLFYQCSNGYTFEQRCEGLYFNYPYQRCDSANVEC 303
Db 418 -----PPTTKPSPTTPK- 431
QY 304 DGEISAPPVTEGNEDEDIDGLDNGCPANFEIDWLLPHGNRCCKYYQCVHGNLVERR 363
Db 431 ---EPAPTTT----- 438
QY 364 CGAGTHFSFELQOCDHIELVGCTLPGESEEVDDVDEDA-CTGWYCTPEIWEPLPNCGP 422
Db 438 -----KSAPTTTKEPAPTTTTPKAAAPNT--PKEPAPTTTKEPAPTTTKEP 470
QY 423 ADFSIDHLLPHESDCGYLQCVHGQTIARPCGNLHFSFATQSCSPVTAGCQVFECDSD 482
Db 471 APPT-----PKEP-----APTTPKAPPT----- 491
QY 483 NOCTSTAAPTAAPTAAPTAAPTAA--APSTVVPPATPPATAAPVPPPT---AIPPT-- 536
Db 491 ---PKEPAPTTTKEPAPTTTTPKAPAPKAPAPTTTPKKTATTPPKLTTTPEKLAPTT 547
QY 536 ---PAPTAAPTAAPTAAPTTPVTP-PTAAPTAAPTTAVPEIPTVT---SAPTAAPT 589
Db 548 EKPAPTPEELAPTTPEEPTPTTPEEPAPTTTPKAAAPNT--PKEPAPTTTKEPAPTTTKEP 605
QY 590 AAPTAAPTAAPTAVPEIPTTTSPTTAAPTAA--APNTT-----VTVPP 634
Db 606 PAPTTTKEPAPTTTTPKGTATTLKEPAPTTTPKPAKELAPTTTKEPTSTSDKPAPTTPK 665
QY 635 TAAPT--AAPAPNTTVVPPPTAAPTAAPTVAH-----APNTTAAVPTTTS 679
Db 666 GTAPTTKEPAPTTTKEPAPTTTPKGTATTLKEPAPTTTPKPAKELAPTTTKEPTSTTS 725
QY 680 ---APATTPEDDDIDP--PLNDPIMP 701
Db 726 DKPAPTTTKEPAPTTTPKAPTTTPKKP 752
```

RESULT 15



US-07-757-022B-52  
Sequence 52, Application US/07757022B  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 CambridgePark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cserr, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 52:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1363 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-52

Query Match 12.0%; Score 520.5; DB 2; Length 1363;  
Best Local Similarity 29.5%; Pred. No. 6.9e-26;  
Matches 203; Conservative 26; Mismatches 205; Indels 253; Gaps 32;  
QY 93 PGPAE-TTQAPATT---QAPTTTQAPTTT-----QAPTTTQ-APTTTQAP 136  
DB 290 PTPKATTTKGPAITPKETPTTPKEPASTTKEPTTIKSAPTTPKEPAPTTKSAP 349  
QY 137 TTTO--APTTTQ--APTTTQ--APTTTQ--APTTTQ--APTTTQ--APTTTQ-- 186  
DB 350 TTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 409  
QY 186 -APTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 243  
DB 410 PAPTTPKEPTTPPKAPPTTPKEPAPTTAPKAPPTTPKEPAPTT-PKEPA-- 467  
QY 244 ELLPNCPADFDIHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNFPYQRCDSANVEC 303  
DB 467 -----PTTTPKEPSPTPK- 480

QY 304 DGEISPAAPVTEGNEDEDIDIGDLLDNGCPANFEIDWLLPHGNRCDDKYYQCVHGNLVERR 363  
DB 480 -----EPAPTTT----- 487  
QY 364 CGAGTHFSFELQCCDHIELVGCTLPGESEEVYDVEDA-CTGMVCTPEPIEWELPNGCP 422  
DB 487 -----KSAPTTTKEPAPTTTTSAPTTTPKEPSPTTTPKE 519  
QY 423 ADFSIDHLLPHESDCGOYLQCVHGQTIARPCPGNLHFSPATQCSFVPTAGCQVFECDSD 482  
DB 520 APTT-----PKEP-----APTTPKKPAPTT----- 540  
QY 483 NOCTSTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 536  
DB 540 ---PKEPAPTTTPKEPAPTTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 596  
QY 536 --PAPTAAAPTAAAPESPTTTPV-PTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 589  
DB 597 EKPAPTTPEELAPTTPEEPTPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTT 654  
QY 590 AAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 634  
DB 655 PAPTTPKETAPTTTPKGAPTTLKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 714  
QY 635 TAAPT-TAAPNTTTPVPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 679  
DB 715 GTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 774  
QY 680 ---APATPEDDDIDP--PLPNDPINP 701  
DB 775 DKPAPTTTPKETAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 801

Search completed: July 18, 1999, 03:46:46  
Job time: 1330 sec

Result No.	Score	Query		Length	DB	ID	Description
		Match	%				
1	313	7.2	907	3	PCF-US95-.04611A-19		Sequence 19, Appl
2	308.5	7.1	451	1	US-08-287-001A-2		Sequence 2, Appl
3	308.5	7.1	451	3	PCF-US95-.09941-2		Sequence 2, Appl
4	379	6.4	2476	2	US-08-276-967-2		Sequence 4, Appl
5	265.5	6.1	214	1	US-08-211-327-4		Sequence 4, Appl
6	258	6.0	2035	2	US-08-479-537A-2		Sequence 2, Appl
7	258	6.0	1867	2	US-08-479-537A-5		Sequence 5, Appl
8	237	5.5	486	1	US-08-450-360-2		Sequence 2, Appl
9	235.5	5.4	1537	1	US-08-325-267A-2		Sequence 2, Appl
10	219.5	5.1	629	3	PCF-US95-.07721-6		Sequence 6, Appl
11	218	5.0	722	1	US-08-347-718B-1		Sequence 1, Appl
12	218	5.0	742	1	US-08-347-718B-2		Sequence 1, Appl
13	218	5.0	742	2	US-08-445-050-2		Sequence 2, Appl
14	218	5.0	722	2	US-08-445-050-3		Sequence 3, Appl
15	218	5.0	722	2	US-08-445-050-7		Sequence 7, Appl
16	218	5.0	722	2	US-08-483-262-1		Sequence 1, Appl
17	218	5.0	742	2	US-08-483-262-2		Sequence 2, Appl
18	218	5.0	745	2	US-08-204-691-2		Sequence 2, Appl
19	218	5.0	722	2	US-08-204-691-3		Sequence 3, Appl
20	218	5.0	722	2	US-08-204-691-7		Sequence 7, Appl
21	217.5	5.0	357	1	US-08-078-683A-8		Sequence 8, Appl
22	216	5.0	2035	1	US-08-046-585-5		Sequence 5, Appl
23	216	5.0	2035	1	US-08-393-703-5		Sequence 5, Appl
24	216	5.0	2035	3	PCF-US93-11721-5		Sequence 5, Appl
25	213	4.9	447	1	US-08-450-360-4		Sequence 4, Appl
26	211	4.9	895	1	US-08-123-161A-8		Sequence 8, Appl
27	211	4.9	895	1	US-08-483-278-8		Sequence 8, Appl
28	206.5	4.8	806	1	US-08-270-076A-11		Sequence 11, Appl
29	206.5	4.8	985	3	PCF-US96-.03916-6		Sequence 6, Appl
30	206.5	4.8	985	3	PCF-US96-.03916-66		Sequence 66, Appl
31	205	4.7	862	1	US-08-325-267A-4		Sequence 4, Appl
32	199	4.6	1848	3	PCF-US95-.10661A-6		Sequence 6, Appl
33	197.5	4.6	921	1	US-08-188-582-2		Sequence 2, Appl
34	197.5	4.6	921	1	US-08-646-715-2		Sequence 2, Appl
35	195.5	4.5	1706	2	US-08-459-568-2		Sequence 2, Appl
36	195.5	4.5	1706	2	US-08-399-411-2		Sequence 2, Appl
37	192.5	4.4	2441	2	US-08-194-468-2		Sequence 2, Appl
38	192.5	4.4	1719	2	US-08-459-568-4		Sequence 4, Appl
39	192.5	4.4	412	2	US-08-477-254A-4		Sequence 4, Appl

Db 528 TTP---TPNATSPAVTTP---TPNATSPGLKTSPTSVAITP-----TPN----- 569  
QY 257 HLLIPHDKYNLFQCSNGYTFEQRCEGLYFNXYVQRDSPANVECGEISAPPPVTEG 316  
Db 569 -----ATSP-----ATSP-----GKTSPTSVAITP--- 585  
QY 317 NEDEDIDGDLNCGPANFEIDWLLPHGNCRCCKYQCVHGNLVERRCGAGTHFSFELQQ 376  
Db 585 -----TPN-----TPN----- 590  
QY 377 CDHIELVGCITLPGGESEEDVDDEACTGWYCTEPIEMEPNPGCPADFSIDHLLPHESD 436  
Db 590 -----ATSP-----TLGKTSPTSVAITP---TPNAT----- 613  
QY 437 CGOYLQCVHGOTIARPCGNLHFSPATOSCESPVTAGCVFECDSNOCSTAAAPTAAPT 496  
Db 613 -----GPIVGETSP-----QANATNHTLGCTSPVPT 640  
QY 497 AAPTAAPTAA-----PTAAPSTVVPVPPATPATAAPVPPPTAAPT--- 536  
Db 641 SQPKNATSAVTTGQHNITSSSTSSMSLRSPNPETLSPSTSDNSTSHMPLTSAHPTGGE 700  
QY 536 ----PAPTAAPTAAPTAAPESTTVTPPTAAPTAAPTAVP-EIPIVTSAP-----T 585  
Db 701 NITQVTPASITHVSTSGE-PRPGTTSOAGSGNSSTSTRPGEVNVTKGTPPONATSP 759  
QY 586 AAPTAAPTAAPTAAPTAVPEIPT-----TVTSPPT---AAPTAAAPAPNTTVT 631  
Db 760 QAPSGOKTAVPTVSTGGKANSTTGKHTTGHGARTSTETPTDYGDSITPRPRYNATY 819  
QY 632 VPP-----TAAPTAAAPNTTVTPVPPPTAAAPTAAAPTVAHAPNTAAAPT--- 677  
Db 820 LPPSTSSKLRPRWTTSPTVTTAAQ---TVVPPTSQPRFNSMLVLQWASLAVLTLL 876  
QY 677 -----TTSAPATTPEDDDID 691  
Db 877 LLVMADCAFRNLSTSHYTTTPPYDDAE 904

RESULT 2  
US-08-287-001A-2  
; Sequence 2, Application US/08287001A  
; Patent No. 5622861  
; GENERAL INFORMATION:  
; APPLICANT: KAPLAN, GERARDO  
; TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
; STREET: Suite 1200, The Candler Bldg, 127 Peachtree  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08287,001A  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spratt, Gwedolyn D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 1414.621  
; TELEPHONE: 404/688-0770  
; TELEFAX: 404/688-9880

; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 451 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-287-001A-2

Query Match 7.1%; Score 308.5; DB 1: Length 451;  
Best Local Similarity 41.6%; Pred. No. 4.3e-14;  
Matches 87; Conservative 4; Mismatches 53; Indels 65; Gaps 10:

QY 94 GPPAET-----TOAPATTAQPTTQAPTPTTQAPTPTTQAPTPTTQAPTPTT 139  
Db 125 GPPRVTTIVRTSTVPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTT 183  
QY 140 QAPTPTTQAPTPTTQAPTPTT---TQAPTPTTQAPTPTT---QAPTPTTQAPTPTT 191  
Db 184 TVPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPT 242  
QY 192 APTTITQAAATPAATT-----PAATT-PAATT-PAATT-----PAAT 226  
Db 243 LPPTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPT 302  
QY 227 TPGVP-----APTSAPVW 239  
Db 303 SPSSQPQAEHPVTLTGATRTQPTSSPLY 331

RESULT 3  
PCT-US95-09941-2  
; Sequence 2, Application PC/TUS9509941  
; GENERAL INFORMATION:  
; APPLICANT:  
; TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS  
; TITLE OF INVENTION: OF USE  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NEEDLE & ROSENBERG, P.C.  
; STREET: Suite 1200, The Candler Bldg, 127 Peachtree  
; STREET: Street  
; CITY: Atlanta  
; STATE: Georgia  
; COUNTRY: USA  
; ZIP: 30303  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent in Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/09941  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: 08/287,001  
; FILING DATE: 5 AUG 1994  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Spratt, Gwedolyn D.  
; REGISTRATION NUMBER: 36,016  
; REFERENCE/DOCKET NUMBER: 1414.621  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 404/688-0770  
; TELEFAX: 404/688-9880  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 451 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; PCT-US95-09941-2





```

Db 184 XRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPAD 243
Qy 239 WPIICELLPNGCPADFDIHLIPHDYCNLFYQCSNGYTFEQRCEGLYFNPYVQRCDS 298
Db 244 XRPX-----PGSTAPX-----AHGVT-----SAP 262
Qy 299 ANVECDGELSPAPVTEGNEDEDDIDIGLLDNGCPANFEIDWLLPHGNKCDKYYOCVIGN 358
Db 263 DXRPXPGSTAPX-----AHG-----278
Qy 359 LVERRCGAGTHFEFLQOCDHIELYGLCTLPGESEEDVDDEACTGWYCP-----TEPIE 413
Db 278 -----VTSAPDXRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPAD 323
Qy 414 WEPLPENG-CPADFSIDHLLPHESDCGOYLQCVHGOTIA----RCPGPNLHFSPATQSCESP 469
Db 324 XRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGS-----TAP 373
Qy 470 VTAGQOVFECDNCTSTAAPTA-APTAAPTAP-----TAAPTAAPSTVVPATP-PATA 524
Db 374 XAHG-----VTSAPDXRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGST 431
Qy 525 AP-VPTTAIP--TPAP-TAAPTAAPTTAAPESPVTIVPPTAAPTAAPTTAAPTAAPTAAP 579
Db 432 APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPX-----PGSTAPXAHGVTSPADXPX 488
Qy 580 VTSAP-----TAAPTAAP-----TAAPTAAPTTAAPTAAPTAAPTAAPTAAPTAAP 626
Db 489 GSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPX 548
Qy 627 NTT-----VTVPTAAPTAAAPNT-TVTVPTAAPT-----TAAPPT--VAHAPNTTAAP 674
Db 549 GSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPX 608
Qy 675 VTT-----TSAPATTPEDDIDP 692
Db 609 GSTAPXAHGVTSPADXPXPGSTAP 633

```

RESULT

```

US-08-479-537A-5
; Sequence 5, Application US/08479537A
; Patent No. 5861381
; GENERAL INFORMATION:
; APPLICANT: CHAMBER, Pierre
; APPLICANT: KIENY, Marie-Paule
; APPLICANT: LATHE, Richard
; APPLICANT: HAREUVENI, Mara
; TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
; TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
; STREET: P.O. Box 1404
; CITY: Alexandria
; STATE: Virginia
; COUNTRY: United States
; ZIP: 22313-1404
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/479,537A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 90/13101
; FILING DATE: 23-OCT-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/FR91/00835

```

```

; FILING DATE: 23-OCT-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/039,320
; FILING DATE: 04-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/403,576
; FILING DATE: 14-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Teskin, Robin L.
; REGISTRATION NUMBER: 35,030
; REFERENCE/DOCKET NUMBER: 017753-025
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 836-6620
; TELEFAX: (703) 836-2021
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1867 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 128..1727
; OTHER INFORMATION: /note= "The amino acids spanning
; OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repea
; OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
; OTHER INFORMATION: repeats varies from 1 to 40."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 134
; OTHER INFORMATION: /note= "Amino acid 134 is X1 = xaa
; OTHER INFORMATION: which is the codon for pro or ala wherein pro = CCT, CCC,
; OTHER INFORMATION: or CCG; and ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 144
; OTHER INFORMATION: /note= "Amino acid 144 is Y = xaa
; OTHER INFORMATION: which is the codon for thr or asn wherein thr = ACT, ACC,
; OTHER INFORMATION: or ACG; and asn = AAT or AAC."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 147
; OTHER INFORMATION: /note= "Amino acid 147 is X2 = xaa
; OTHER INFORMATION: which is the codon for pro or ala wherein pro = CCT, CCC,
; OTHER INFORMATION: or CCG; and ala = GCT, GCC, GCA, or GCG."
; FEATURE:
; NAME/KEY: Peptide
; LOCATION: 1..21
; OTHER INFORMATION: /note= "Amino acids 1 to 21 are a
; OTHER INFORMATION: 21 amino acid precursor sequence."
; US-08-479-537A-5

```

```

Query Match 6.0%; Score 258; DB 2; Length 1867;
Best Local Similarity 23.4%; Pred. No. 6e-10;
Matches 160; Conservative 43; Mismatches 280; Indels 202; Gaps 34;

Qy 93 PGPAETTQAPATTQAPTTTQAPTT-----TTQAPTTTQAPTTTQAPTTTQAPTTT 145
Db 66 PGSGSTTQGDVTLAPATEPAGSAAATWGQDVTSVPVTRPALGSTTTPAHADVTSAPDNK 125
Qy 146 QAPTTTQAPTT--TQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTT 198
Db 126 PAPGST-APXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGST-APXAHGVTSPAD 183
Qy 199 AATTAATTAA-----TTTAATTAA-----TTTAATTAA-----TTTAATTAA 238
Db 184 XRPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPADXPXPGSTAPXAHGVTSPAD 243
Qy 239 WPIICELLPNGCPADFDIHLIPHDYCNLFYQCSNGYTFEQRCEGLYFNPYVQRCDS 298
Db 244 XRPX-----PGSTAPX-----AHGVT-----SAP 262

```

QY	299	ANVECDGEISAPPVTVTSGNEDEDIDIGDLDNGCPANFEIDMLLPHGNRCDDKYIQCYPHGN	358
DB	263	DXXRXPGSTAPX-----AHG-278	
QY	359	LVERRCGAGTHFSFELQQCDHIELVGCTLPGGSEEEVDVEDACTGWICP-----TEPIE	413
DB	278	-----VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPD	323
QY	414	WEPLPNG-CPADEFSIDLLHPHSDCCQYLQCVHGQTIA---RPCGNLHFSPATQSCESP	469
DB	324	XRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGS-----TAP	373
QY	470	VTAGCQVFECDSNQCTISTAFTA-APTAAPTAAAP--TAAPTAAAPTSTVVVPATP-PATA	524
DB	374	XAHG--VTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGST	431
QY	525	AP-VPPPTAIP--TPAP-TAAPTAAPTAAPESSPTTVTPPTAAPTAAPTTAVPEI-PIT	579
DB	432	APXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX--PGSTAPXAHGVTSAPDXRPX	488
QY	580	VTSAP-----TAAPTAAP--TAAPTAAPTAVPEI-PTTVTSPPTAAPTTAA----PAP	626
DB	489	GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX	548
QY	627	NTT-----TVTPPTAAPTAAAPNT-TVTPPTAAP--TAAPT--VAHANPTAAPT	674
DB	549	GSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPXPGSTAPXAHGVTSAPDXRPX	608
QY	675	VTT-----TSAPATTPEDDDDIDP	692
DB	609	GSTAPXAHGVTSAPDXRPXPGSTAP	633

RESULT 9  
US-08-325-267A-2  
; Sequence 2, Application US/08325267A  
; Patent No. 5585271  
; GENERAL INFORMATION:  
; APPLICANT: WATARI, JUNJI  
; APPLICANT: TAKATA, YOSHIHIRO  
; APPLICANT: OGAWA, MASASHIRO  
; APPLICANT: PENNILLA, MERJA  
; APPLICANT: ONNELA, MAIJA-LEENA  
; APPLICANT: KERANEN, SIRKA  
; TITLE OF INVENTION: YEAST AGGLUTINATION GENES AND YEAST  
; TITLE OF INVENTION: CONTAINING THEM  
; NUMBER OF SEQUENCES: 7  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT  
; STREET: 1755 S. JEFFERSON DAVIS HWY, SUITE 400  
; CITY: ARLINGTON  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22202  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/325,267A  
; FILING DATE: 18-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP PCT/JP94/00290  
; FILING DATE: 24-FEB-1994  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 38871/1993  
; FILING DATE: 26-FEB-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: OBLON, NORMAN F  
; REGISTRATION NUMBER: 24,618  
; REFERENCE/DOCKET NUMBER: 2589-023-0XPCT



```

:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 703-413-3000
:
: TELEFAX: 703-413-2220
:
: TELEX: 248955 OPAT UR
:
: INFORMATION FOR SEQ ID NO: 2:
:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 1537 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08325-267A-2

```

Query Match: 5.4%; Score 235.5; DB 1; Length 1537;  
Best Local Similarity 23.2%; Pred. No. 1.6e-08;  
Matches 166; Conservative 45; Mismatches 291; Indels 213;

[illegible]

```

RESULT 10 ;
PCT-US95-07721-6 ;
; Sequence 6, Application PC/TUS9507721 ;
; GENERAL INFORMATION: ;
; APPLICANT: Massachusetts Institute of Technology ;
; TITLE OF INVENTION: Class BI and CI Scavenger Receptors ;
; NUMBER OF SEQUENCES: 8 ;

```

CORRESPONDENCE ADDRESS:  
ADDRESSEE: Patrea L. Pabst  
STREET: 2800 One Atlantic Center  
STREET: 1201 West Peachtree Street  
CITY: Atlanta  
STATE: Georgia  
COUNTRY: USA  
ZIP: 30309-3450  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/07721  
FILING DATE:  
CLASSIFICATION:  
ATTORNEY/AGENT INFORMATION:  
NAME: Pabst, Patrea L.  
REGISTRATION NUMBER: 31,284  
REFERENCE/DOCKET NUMBER: MIT6620  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (404) 873-8794  
TELEFAX: (404) 873-8795  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 629 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO  
FRAGMENT TYPE: internal  
FEATURE:  
NAME/KEY: misc. feature  
LOCATION: 1..629  
OTHER INFORMATION: /Function = "Amino acid sequence for the  
OTHER INFORMATION: Drosophila Melanogaster Scavenger Receptor  
OTHER INFORMATION: Class CI."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 30..353  
OTHER INFORMATION: /note= "Positions 30-32, 90-92,  
OTHER INFORMATION: 129-131, 180-182, 253-255 and 351-353  
OTHER INFORMATION: represent potential N-glycosylation  
OTHER INFORMATION: sites."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 1..20  
OTHER INFORMATION: /note= "Amino acids 1-20 represent  
OTHER INFORMATION: a putative signal sequence."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 21..74  
OTHER INFORMATION: /note= "Amino acids 21-74 represent  
OTHER INFORMATION: complement control protein domain  
OTHER INFORMATION: number 1."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 75..127  
OTHER INFORMATION: /note= "Amino acids 75-127  
OTHER INFORMATION: represent complement control protein  
OTHER INFORMATION: domain number 2."  
FEATURE:  
NAME/KEY: Modified-site  
LOCATION: 128..312  
OTHER INFORMATION: /note= "Amino acids 128-312  
OTHER INFORMATION: represent an MAM domain."  
FEATURE:  
NAME/KEY: Disulfide-bond  
LOCATION: 22..381  
OTHER INFORMATION: /note= "The cysteines at positions  
OTHER INFORMATION: 22, 45, 59, 72, 77, 99, 113, 125, 136, 144  
OTHER INFORMATION: 217, 254, 310, 339, 343, 361, 363, 367, 371  
OTHER INFORMATION: and 381 form disulfide bonds."



```

; GENERAL INFORMATION:
; APPLICANT: Wang, Chu-San
; APPLICANT: Tang, Jordan J.N
; TITLE OF INVENTION: METHOD FOR REDUCING INTESTINAL ABSORPTION OF
; TITLE OF INVENTION: CHOLESTEROL
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Patrea L. Pabst
; STREET: 2800 One Atlantic Center
; STREET: 1201 West Peachtree Street
; CITY: Atlanta
; STATE: Georgia
; COUNTRY: USA
; ZIP: 30309-3450
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/347,718B
; FILING DATE: December 1, 1994
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Pabst, Patrea L.
; REGISTRATION NUMBER: 31,284
; REFERENCE/DOCKET NUMBER: OMR#150
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (404) 873-8794
; TELEFAX: (404) 873-8795
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 742 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; FRAGMENT TYPE: internal
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 186..187
; OTHER INFORMATION: /note= "Position 187 represents a
; OTHER INFORMATION: potential N-linked glycosylation site."
; FEATURE:
; NAME/KEY: Modified-site
; LOCATION: 193..194
; OTHER INFORMATION: /note= "The serine at position 194
; OTHER INFORMATION: represents an active site serine."
; FEATURE:
; NAME/KEY: misc. feature
; LOCATION: 1..742
; OTHER INFORMATION: /Function = "Amino acid sequence for
; OTHER INFORMATION: the Human Milk Bile Salt-activated Lipase."
;
US-08-347-718B-2

Query Match          5.0%; Score 218; DB 1; Length 742;
Best Local Similarity 24.9%; Pred. No. 1.1e-07;
Matches 117; Conservative 44; Mismatches 188; Indels 120; Gaps 23;

QY 304 DGEISAPPVTGNEDEDIDI---GDLLNGCPANFEIDLWLLPHGNRCDK-----YQC 354
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 312 DGDIFPADINLYANAADIYAGTNNDMGHIFASID---MPAINKGNKKVTEEDFYKL 367
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 355 VHGNLVERC-CAGTHFSELOQCCHIELVGTLPGESEEE---VDVDEDACTGWCPTPE 410
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 368 VSEFITKGLRKAKTTFDVTYES-----WAQDPSQENKKTVDVDFDVL--FLVPPE 418
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 411 PIEWEPLPNCPADFSIDLHLLPHES-----DCGQYLQCVHGQTIARPCPCNLHF 459
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 419 IALAQRANAKSAK-TYALFSLHPSEMPYYPKWVGADHADDIQQYVGKPFATP-----TCY 473
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 460 SPATQSCSPVTAGCQVFECDSNOCSTSTAAPT----- 493
```

```

Db 474 RPQDRTVSKAMIAWTNFAKTGDPNMGDSAVPTHPEPTTENSGLYLEITKKMGSSSMKRS 533
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 493 -----AAPTAAPTAAPTAAPTA-APSTVVVPPATPPATAAPVPPTTAIPTPA 537
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 534 LRTNFRYWTLYLALPTVTDQEATVPPTGDSSEATVPP-TGDSEATAPVPT-----G 586
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 538 PTAAPTAAPTTAAPESPSTTVVPPPT---AAPTAAPT--TAVPEIPITVTSAPTAAPTAAP 592
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 587 DSGAPPVPPTGDSGAPP-----VPPTGDSGAPPVPPTGDSGAPPVPPTGDSGAPPVPPTGD 642
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 TAAPTAAPT---TAVPEIPTVTSPTTAAPTAAAPTAAAPTNTVTVPT---AAPTAAAPT 646
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 643 SGAPPVPPTGDSGAPPVPPTGDPAGPPVPPTGDSGAP-----PVPTGDSGAPPVPTPTGDS 698
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 647 TTVTVPTTAAPTAAAPTVAHAPNTTAAAPTVTTSAPATTPEDDDDIDPPLP 695
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 699 ETAPVPPT-GDSGAPP-VPTGDSSEAPV-----PPTDSDSKEAOMP 737
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 13
US-08-445-050-2
; Sequence 2, Application US/08445050
; Patent No. 5763739
; GENERAL INFORMATION:
; APPLICANT: Blaeckberg, Lars
; APPLICANT: Edlund, Michael
; APPLICANT: Hansson, Lennart
; APPLICANT: Herneli, Oille
; APPLICANT: Lundberg, Lennart
; APPLICANT: Stroemqvist, Mats
; APPLICANT: Toernell, Jan
; TITLE OF INVENTION: NO. 5763739el Polypeptides
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: White & Case
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States
; ZIP: 10036-2787
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/445,050
; FILING DATE:
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/204,691
; FILING DATE:
; APPLICATION NUMBER: SE 9300686-4
; FILING DATE: 01-MAR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: SE 9300722-7
; FILING DATE: 04-MAR-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Sterner Ph.D., Richard J
; REGISTRATION NUMBER: 35,372
; REFERENCE/DOCKET NUMBER: 1103326-850
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)819-8783
; TELEFAX: (212)354-8113
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 745 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-445-050-2
```

Query Match 5.0%; Score 218; DB 2; Length 745;  
Best Local Similarity 24.9%; Pred. No. 1.le-07;  
Matches 117; Conservative 44; Mismatches 188; Indels 120; Gaps 23;

QY 304 DGEISPAPVTEGNEDEDIDI---GDLIDGCPANFEIDWLLPHGNRCDK-----YYOC 354  
DB 315 DGFDPADPINLYANAADIDYAGTNMMDGHIFASID---MPAINKNKKVTEEDFYKL 370  
QY 355 VHGNLVERRC-GAGTHFSFELQOCDHIELVGTCLPGSEEE---VDVDEDACTGWYCPTE 410  
DB 371 VSEFTITKGLRGAKTTFDYYTES-----WAQDPQENKKKTVDVDFEVDVL--FLVPTE 421  
QY 411 PWEPLPNCPCPADFSIDHLLPHES-----DCQOYLQCVHGQTIARPCPNLHF 459  
DB 422 IALAQRANAKSAK-TYAYLFSHPSRMPVYKMWGADHADDIQQYVFGKPFATP---TCY 476  
QY 460 SPATQSCSPVTTAGCOVFECDSNQCSTAAPT-----PPTDSDSKEAOMP 493  
DB 477 RPQDRTVSKAMIAWTNFAKTGDPNMGDSAVPTHWEPTTENSGLYLEITKKMGSSMKRS 536  
QY 493 -----AAPTAAPTAAPTAAPTA-APSTVVPVPPATPPATAAPVPPPTTAIPTPA 537  
DB 537 LRTNELYWTLTYLALPTVTDQEATVPVPTGSEATPVP--TGDSETAPVPT-----G 589  
QY 538 PTAAPTAAPTTAAPESPTTVTPPT---AAPTAAPT--TAVPEIPTVTSAPTAAPTAAP 592  
DB 590 DSGAPPVPTGDSGAPP-----VPPTGDSGAPPVPTGDSGAPPVPTGDSGAPPVPTGD 645  
QY 593 TAAPTAAPT--TAVPEIPTVTSAPTAAPTTAAPNTTAAPTTTPEDDDIDPPLP 695  
DB 646 SGAPPVPTGDSGAPPVPTGAGPPVPVPTGDSGAP-----PVPPTGDSGAPPVPTGDS 701  
QY 647 TTVTVPTTAAPTAAPTVAHAPNTTAAPTTTSAPTATTPEDDDIDPPLP 695  
DB 702 ETAPVPT--GDSGAPP-VPTGDSAPV-----PPTDSDSKEAOMP 740

RESULT 14  
US-08-445-050-3  
; Sequence 3, Application US/08445050  
; Patent No. 5763739  
; GENERAL INFORMATION:  
; APPLICANT: Blaeckberg, Lars  
; APPLICANT: Edlund, Michael  
; APPLICANT: Hansson, Lennart  
; APPLICANT: Hernell, Olle  
; APPLICANT: Lundberg, Lennart  
; APPLICANT: Stroemqvist, Mats  
; APPLICANT: Toernell, Jan  
; TITLE OF INVENTION: No. 5763739el Polypeptides  
; NUMBER OF SEQUENCES: 21  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: White & Case  
; STREET: 1155 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: United States  
; ZIP: 10036-2787  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/445,050  
; FILING DATE:  
; CLASSIFICATION: 800  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/08/204,691  
; FILING DATE:  
; APPLICATION NUMBER: SE 9300686-4  
; FILING DATE: 01-MAR-1993

PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: SE 9300722-7  
; FILING DATE: 04-MAR-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Sterner Ph.D., Richard J  
; REGISTRATION NUMBER: 35,372  
; REFERENCE/DOCKET NUMBER: 1103326-850  
; TELEPHONE: (212)819-8783  
; TELEFAX: (212)354-8113  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 722 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens  
; TISSUE TYPE: Mammary gland  
US-08-445-050-3

Query Match 5.0%; Score 218; DB 2; Length 722;  
Best Local Similarity 24.9%; Pred. No. 1.le-07;  
Matches 117; Conservative 44; Mismatches 188; Indels 120; Gaps 23;

QY 304 DGEISPAPVTEGNEDEDIDI---GDLIDGCPANFEIDWLLPHGNRCDK-----YYOC 354  
DB 292 DGFDPADPINLYANAADIDYAGTNMMDGHIFASID---MPAINKNKKVTEEDFYKL 347  
QY 355 VHGNLVERRC-GAGTHFSFELQOCDHIELVGTCLPGSEEE---VDVDEDACTGWYCPTE 410  
DB 348 VSEFTITKGLRGAKTTFDYYTES-----WAQDPQENKKKTVDVDFEVDVL--FLVPTE 398  
QY 411 PWEPLPNCPCPADFSIDHLLPHES-----DCQOYLQCVHGQTIARPCPNLHF 459  
DB 399 IALAQRANAKSAK-TYAYLFSHPSRMPVYKMWGADHADDIQQYVFGKPFATP---TCY 453  
QY 460 SPATQSCSPVTTAGCOVFECDSNQCSTAAPT-----PPTDSDSKEAOMP 493  
DB 454 RPQDRTVSKAMIAWTNFAKTGDPNMGDSAVPTHWEPTTENSGLYLEITKKMGSSMKRS 513  
QY 493 -----AAPTAAPTAAPTAAPTA-APSTVVPVPPATPPATAAPVPPPTTAIPTPA 537  
DB 514 LRTNELYWTLTYLALPTVTDQEATVPVPTGSEATPVP--TGDSETAPVPT-----G 566  
QY 538 PTAAPTAAPTTAAPESPTTVTPPT---AAPTAAPT--TAVPEIPTVTSAPTAAPTAAP 592  
DB 567 DSGAPPVPTGDSGAPP-----VPPTGDSGAPPVPTGDSGAPPVPTGDSGAPPVPTGD 622  
QY 593 TAAPTAAPT--TAVPEIPTVTSAPTAAPTTAAPNTTAAPTTTPEDDDIDPPLP 695  
DB 623 SGAPPVPTGDSGAPPVPTGAGPPVPVPTGDSGAP-----PVPPTGDSGAPPVPTGDS 678  
QY 647 TTVTVPTTAAPTAAPTVAHAPNTTAAPTTTSAPTATTPEDDDIDPPLP 695  
DB 679 ETAPVPT--GDSGAPP-VPTGDSAPV-----PPTDSDSKEAOMP 717

RESULT 15  
US-08-445-050-7  
; Sequence 7, Application US/08445050  
; Patent No. 5763739  
; GENERAL INFORMATION:  
; APPLICANT: Blaeckberg, Lars  
; APPLICANT: Edlund, Michael  
; APPLICANT: Hansson, Lennart  
; APPLICANT: Hernell, Olle  
; APPLICANT: Lundberg, Lennart  
; APPLICANT: Stroemqvist, Mats  
; APPLICANT: Toernell, Jan  
; TITLE OF INVENTION: No. 5763739el Polypeptides



**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK (USPTO)**







Db	349	FCRCVCSGEVLGNLNSRPS-RSLPTT-----TTPALPAPLPPPLPPLPLP	394
Qy	496	--TAAP-----TAAPTAATAAP-----STVVVPPA-----TPATAAP-----	527
Db	395	INTAVPIPPPLPVYALAPPLPLALPLISGVVPPAPPYPPGKWTTPPLAPAPPEKIV	454
Qy	527	--VPTTAIP-----TPAPTAAPTAAPTAAPESPTTVTVVPT-----AAPTAA	568
Db	455	PVLPPGSCPPSEKPNPAPPEPEPKSSPALPPAPPAPSMPSAVRVPSPPIPPAPPA	514
Qy	569	PTTAVPELPIVTVISAPTA-----APTAAPT-----AAPTAAPTAVP	605
Db	515	PRASMPALPPAPPPSPATRLCPLPPSPAPNPPAPPAPPTPKLLSANPPCPPVPPAP	574
Qy	606	EIPTVTSPTTAAPTAAAPANTTVTPPTAAPTTAAPANTTVTPPTAAPTAAAPTVA	665
Db	575	NRPPAPAPAPPELAPPDPPT-----PPVANSPPAPPAP-----PAPPSALPFVNP	624
Qy	666	HAPNTTAAPTVTTISAP-----ATTPEDDDIDPLPLNDPINP	701
Db	624	APPTPAAPKSRPALPAAPAPPAPPVVRATTPPPAPPAPPAPNSMALP	670

RESULT 4

W43106	W43106 standard; Protein; 1664 AA.
AC	W43106;
DT	16-OCT-1998 (first entry)
DE	C. thermocellum OlpB protein.
KW	Multimer; enzyme; complex; protein-protein interaction; dockerin domain;
KW	cohesin domain; catalytic subunit; scaffold subunit; SdBA; synergistic;
KW	cellulosome integrating protein; scaffoldin dockerin binding protein.
OS	Clostridium thermocellum.
FH	Key Location/Qualifiers
FT	Domain 28..192
FT	/note="cohesin type II domain"
FT	Domain 207..363
FT	/note="cohesin type II domain"
FT	Domain 409..565
FT	/note="cohesin type II domain"
FT	Domain 607..763
FT	/note="cohesin type II domain"
PN	FR2748479-A1.
PD	14-NOV-1997.
PF	10-MAY-1996; 005854.
PR	10-MAY-1996; FR-005854.
PA	(INSP ) INST PASTEUR.
PI	Beguin P, Leibovitz E;
DR	DR - WPI; 98-011569/02.
DR	N-PSDB; T86623.
PT	Cellulase proteins with cohesin or dockerin type II domains - useful
PT	for potentiating the activity of multiprotein enzyme complexes
PS	Claim 7; Page 31-39; 60pp; French.
CC	Multimeric protein, especially enzymatic, complexes are held together
CC	by protein-protein interactions between domains designated dockerins
CC	and cohesins, which are found on the catalytic and scaffold subunits
CC	respectively. An example of such a complex is the cellulose degrading
CC	protein complex from Clostridium thermocellum, known as the cellulosome.
CC	This complex comprises around 15 proteins including endoglucanases,
CC	cellulohydrolases, hemicellulases, e.g. xylanases or lichenases, which
CC	interact with a central "scaffold" protein designated the cellulosome
CC	integrating protein (Cipa; see W43108). The catalytic subunits interact
CC	with the Cipa subunit via conserved 23 amino acid dockerin domains. Cipa
CC	has been shown to contain 9 copies of a cohesin domain.
CC	The invention relates to the isolation of proteins binding to a novel
CC	dockerin type domain found in the C-terminal portion of Cipa. The new
CC	domain is designated a type II dockerin domain (as compared to the type
CC	domain found on the catalytic subunits of the cellulosome). The type II
CC	dockerin domain has some sequence similarity to the type I dockerins but
CC	is unable to bind type I cohesin domains.
CC	The sequence presented here is an example of a protein which binds
CC	the novel type II dockerin domain and is the product of the OlpB gene.
CC	The protein contains 4 type II cohesin domains in the N-terminal portion





QY 641 AAPAPNTVTVDPPT-----AAPTA-----APPTVAHAPNTTAAPVTTTSA 680  
 ID W15761  
 AC W15761 standard; Protein: 297 AA.  
 DT 19-JUN-1997 (first entry)  
 DE Cotton fibrous tissue specific protein KC03.  
 KW Fibrous tissue; cotton fibre; cotton flower; Gossypium barbadense;  
 OS Gossypium hirsutum.  
 PN J09075093-A.  
 PD 25-MAR-1997.  
 PF 20-FEB-1996; 031987.  
 PR 21-FEB-1995; US-391696.  
 PR 29-DEC-1995; US-580545.  
 PA (TOYM ) TOYORO KK. TECH.  
 PA (UYTE-) UNIV TEXAS TECH.  
 DR WPI; 97-239274/22.  
 DR N-PSDB; T60569.  
 PT Cotton fibrous tissue gene - used to produce transformants with  
 improved cotton fibres, and improved yield  
 PS Claim 3; Page 21-22; 3pp; Japanese.  
 CC W15759-W15763 represent cotton fibrous tissue proteins of the invention.  
 CC The genes encoding these sequences are specifically expressed during the  
 CC growth of cotton fibre. The genes were isolated from the polyA RNA of the  
 CC ovules of a cotton flower of Gossypium barbadense, and G. hirsutum. The  
 CC genes encoding these sequences, and vectors containing them are used in  
 CC the preparation of cotton transformants, with improved cotton fibres and  
 CC improved yield.  
 SQ Sequence 297 AA;

Query Match 7.3%; Score 317; DB 1; Length 297;  
 Best Local Similarity 35.8%; Pred. No. 1.2e-10;  
 Matches 101; Conservative 8; Mismatches 99; Indels 74; Gaps 17;  
 QY 497 AAPTAAPTAAPTAAPVPPATAPVPPPTTAIPTAPTAAPTAAPTAAPESPTT 556  
 DB 35 AAP--VPVKAPTA-----PPVKPTT--PVPYKA-PTAP--PTKPTPKP--PTK 79  
 QY 557 VTVPTAAPTAAPTAV-PEIPTVTSAPT-----AAPTAAPT-----AAPTA 599  
 DB 80 APTPPYKPTKAPTTPYKPPAPAPPTKAPTTPYKPPAPAPPTKAPTTPYKPPAPAPPTKA 139  
 QY 600 PTAPEIPTVTSPTAAPT-----TAAPAPNTVTVPPTAAPTAAAPNTVTVPP 653  
 DB 140 PT---PPFKPPAPAPTAPPTTPYKPTPPAPAPVKAAPTTPYMPPTPKAP-TPAPAP 195  
 QY 654 TAAPTAAPTVAHAPNTAAAPTPTTTSAPATTPEDDDDIDPLPNDPINCVECNVLPWAH 713  
 DB 196 TKAPT--PPYKPPVTPPKPTTAPYKPP-----SPPLP--PVR-----TK 235  
 QY 714 ADCDKYWCDDGNQVLLVVCSEGLQFNPTTKTCDFAACNVCYR 755  
 DB 236 KDC-----IPLCGQRCKLHSLNLCRLACMTCCDR 265

RESULT 9  
 P50073  
 ID P50073 standard; Protein: 907 AA.  
 AC P50073;  
 DT 17-OCT-1991 (first entry)  
 DE Epstein-Barr virus (EBV) outer surface protein.  
 KW Epstein-Barr virus; antigen; vaccine.

OS Epstein-Barr virus.  
 PN EP-151079-A.  
 PD 07-AUG-1985.  
 PR 28-JAN-1985; 400141.  
 PR 30-JAN-1984; US-575352.  
 PR 23-JUL-1984; US-633558.  
 PA (UYCH-) UNIV OF CHICAGO.  
 PI Kleff E, Tanner J, Hummel M, Belsel C;  
 DR WPI; 85-191978/32.  
 DR N-PSDB; N50114.  
 PT New fragment of Epstein-Barr Virus DNA - useful in vector to  
 express polypeptide for use in prepn. of vaccine against the  
 PT virus and for use in diagnosis.  
 PS Claim 2; Page 23-25; 26pp; English.  
 CC The sequence encodes an outer surface viral protein of EBV, used  
 CC to generate antibodies reacting with the surface proteins of  
 CC EBV-infected cells, and in the preparation of a vaccine against EBV.  
 SQ Sequence 907 AA;

Query Match 7.2%; Score 313; DB 1; Length 907;  
 Best Local Similarity 22.5%; Pred. No. 6e-10;  
 Matches 168; Conservative 65; Mismatches 221; Indels 294; Gaps 32;  
 QY 43 WLLPHEVDCTKYYCEYGLKFIAPDCAPGTEFKFSAGTCVHAALAGC-TLPGPPAET-- 100  
 DB 352 WAMPNNTETD--FKCKWTLTSGTSPGC-ENISGAFASNRTEITIVSGLTAPKTLIIIRT 408  
 QY 100 -TOAPATOAPTTTOAAPTITTOAAPTITTOA--PTTTOAAPTITTOAAPTITTOAPT 155  
 DB 409 ANNATTTTHKVFISKAPESTTSTPTLNTGADPNTTTLGSPSTHVTNLTAPAST-GPT 467  
 QY 156 TTTQAPTTTTQAPTTTTQAPTT-----TQAP--TTTQAPTTTTQA--TTITQAA 200  
 DB 468 VSTADVTSPTAGTSGASPVTPSPWDNCTESKAPDMTSTSPVTTPTFNATSPPTAV 527  
 QY 201 TTPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVPPICELLNPGCADFDI 256  
 DB 528 TTP--TPNATSPPTAVTTP--TPNATSPTLGKTSPTSAVTP--TPN----- 569  
 QY 257 HLLIPHDKYCNLFYQCSNGYTFEQRCPGLYFNYPVQRCDSPANVECDGEISAPPPVTEG 316  
 DB 569 -----ATSPTL-----GKTSPTSAPT-- 585  
 QY 317 NEDEDIDIGLLDNGCPANFEIDWLLPHGNCRCDKYYQCWGNLVERRCGAGTHFSFELQ 376  
 DB 595 -----TTPN----- 590  
 QY 377 CDHIELVGCTLPGGESEEVODEACTGWYCPTEPIEWEPLNPGCPADFSIDLHPHESD 436  
 DB 590 -----ATSP-----TLGKTSPTSAPT--TPTPNAT----- 613  
 QY 437 CGQYLQCVHGOTIARPCGNLHFSPATQSCSPYTAGOVFECDSDNOCTSTAAPTAPT 496  
 DB 613 -----GPTVGETSP-----OANATNHTLGGTSPVVT 640  
 QY 497 AAPTAAPTAA-----PTAAPSTVVPVPPATPAPPTTAIPT----- 536  
 DB 641 SOPKNATSAVTTGQHNITSSSTSSLSRLSPSNPETLSSTSDNSTSHMLTSHAHTGGE 700  
 QY 536 -----PAPTAAPTAAPTAAPTSPVTVPPTAAPTAAPTAVP-EIPTVTSAP----- 585  
 DB 701 NITQVTPASISTHHVSTSSPE-PRPGTTTQSASGPGNSSTSKPGEVNVTKGTPPQNATSP 759  
 QY 586 AAPTAAPTAAPTAAPTAVPEIPT-----TVTSPT-----AAPTAAPTAAPTT 631  
 DB 760 QAPSGOKTAVETVSTGOKANSTTGGKHTTGHGARTSTEPTTDYGGDSTTTPRPNATY 819  
 QY 632 VPP-----TAAPTAAAPANTTTPVPPTAAPTAAAPTVAHAPNTTAAPT----- 677  
 DB 820 LPPSTSSKLRPRWFTTSPVTTAQA---TVPVPTTSQPRFNSLMLVQLQWASLAVITLL 876  
 QY 677 -----TTSAPATTPEDDDID 691

Db 877 LLVADCAFRRLNSTSHYTPPYDDAE 904  
 :|| ||| ||| :

## RESULT 10

R80144  
 ID R80144 standard; Protein; 907 AA.  
 AC R80144; 1996 (first entry)  
 DE 18-JAN-1996  
 KW EBV gp350; gp220; gp350/gp220; non-splicing variant; vaccine.  
 OS Epstein-Barr virus.  
 FH Key Location/Qualifiers  
 FT Peptide  
 FT 1..18  
 FT /label= Sig\_peptide  
 FT 861..881  
 FT /note= "transmembrane region"  
 PN W09528488-AL.  
 PD 26-OCT-1995.  
 PF 13-APR-1995; U04611.  
 PR 18-APR-1994; US-229291.  
 PA (AVIR-) AVIRON.  
 PI Jackman WT, Spaete R;  
 DR WPI; 95-373802/48.  
 DR N-PSDB; T04821.  
 PT New DNA encoding a homogeneous gp350 protein - can be used for  
 PT preventing and treating Epstein-Barr virus-related diseases or  
 PT conditions  
 PS Disclosure; Fig.1; 61pp; English.  
 CC The donor and acceptor splice sites of the EBV gene encoding gp350/  
 CC 220 are mutated by replacement of native nucleotides by non-native  
 CC nucleotides, without altering the encoded amino acid sequence,  
 CC resulting in elimination of gp220 prodn. Recombinant homogeneous  
 CC gp350, useful in vaccines, is expressed in mammalian or insect cell  
 CC hosts.  
 SQ Sequence 907 AA;

Query Match 7.2%; Score 313; DB 1; Length 907;

Best Local Similarity 22.5%; Pred. No. 6e-10;  
 Matches 168; Conservative 65; Mismatches 221; Indels 294; Gaps 32;

Qy 43 WLLPHEYDCTFEYCEYGLKFIAPRDCAAGTEFEKSAQTCVHAALAGC-TLPGPAAET-- 100  
 Db 352 WAWPNTETD--FKCKWLTSGTPSGC-ENISGAFASNRTEFDITVSGLGATPKTLIIIRT 408  
 Qy 100 -TOAPATTOAPTTOAPTTOAPTTOA---PTTTOAPTTOAPTTOAPTTOAPTTOAPT 155  
 Db 409 ATNATTTTHKVFISKAPSTTSPLNTGTGADPNWTTGLSSHTVNLNAPAST-GPT 467  
 Qy 136 TTTOAPTTOAPTTOAPTTOAPTTOA---TOAP--TTTTOAPTTOAPT--TTITQAA 200  
 Db 468 VSTADVTSTPATGTTSGASPVTPSPWDNGTESKAPDWTSSTSPTTTPNATSPTRAV 527  
 Qy 201 TPAATTPAAT--TPAATTPAATTPAATTP--GVDPATPSAPVWPPICELLNGCPADFDI 256  
 Db 528 TTP--TPNATSPTRAVTTP--TPNATSPTRAVTTP--TPNATSPTRAVTTP--TPN 569  
 Qy 257 HLLPHDKYCNLFYQCSNGYTFEQRCPGLYFNPVVQRCDSANVECDGEISPAAPPVTEG 316  
 Db 569 -----ATSP-----ATSTPL-----GKTSPTSAPT-- 585  
 Qy 317 NEDEDIDIGLLDNGCPANFEIDLPHGNCCKYQCVHGNLVERRGAGTHFSFELQ 376  
 Db 585 -----TPPN----- 590  
 Qy 377 CDHIELVGTCLPGGESEVVDVEDACTGWCPTEIENEPLPNCGPADFSIDLHPHESD 436  
 Db 590 -----ATSP-----TLGKTSPTSAPT--TPNAT-- 613  
 Qy 437 CGOYLQCVHGOTIARPCGNLHFSFATOSCESPVTAGCQVFECDSNOCSTAPTAAPT 496  
 Db 613 -----GPTVGTSP-----QANATNHLGTSPTPVVT 640

Qy 497 AAPTAAPTAA-----PTAAPSTVVPPATPPATAAPVPPPTTAIPT--- 536  
 Db 641 SQPKNATSAVTTGQHNITSSSTSMRLSPSSNPETLSPTSDNSTSHMPLLTSAHPTGGE 700  
 Qy 536 ---PAPTAAPTAAAPESPTTVVPPTAAPTAAAPTAAV-EIPTVTSAP-----T 585  
 Db 701 NITQVTPASISTHVSSTSSPE-PRFGTTSQASGPGNSSTSKPGEVNVTKGTPPONATSP 759  
 Qy 586 AAPTAAPTAAAPTAAAPTAAVPEIPT-----TVTSPTT---AAPTAAPAPNTTPT 631  
 Db 760 QAPSGOKTAVPTVTSKGANSTTGGKHTTGCHGARTSTPTDYGGDSTTTPRPNATY 819  
 Qy 632 VPP-----TAAPTAAAPANTTVVPPTAAPTAAAPTAAVHAAPNTAAAPT--- 677  
 Db 820 LPPSTSSKLRPRWTFSTPPVTTAQA---TVDPVPTSQPRFSNLSMLVQWASLAVLTLL 876  
 Qy 677 -----TTSAPATTPEDDDID 691  
 Db 877 LLVADCAFRRLNSTSHYTPPYDDAE 904

## RESULT 11

R92803  
 ID R92803 standard; Protein; 451 AA.  
 AC R92803;  
 DT 24-MAY-1996 (first entry)  
 DE Hepatitis A virus receptor.  
 KW Hepatitis A virus; receptor; HAV; diagnosis; therapy; vaccine;  
 KW transgenic animal.  
 OS Cercopithecus aethiops.  
 PN W09604376-AL.  
 PD 15-FEB-1996.  
 PF 04-AUG-1995; U09941.  
 PR 05-AUG-1994; US-287001.  
 PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
 PI Feinstein SM, Kaplan G;  
 DR N-PSDB; T17864.  
 DR DNA encoding Hepatitis A virus receptor and its fragments - useful  
 PT for the detection, diagnosis, prevention and treatment of HAV  
 PT infection  
 PS Claim 1; Page 39-41; 52pp; English.  
 CC A hepatitis A virus (HAV) receptor (R92803) was identified as the  
 CC product of a nucleic acid sequence isolated from HAV-infected  
 CC primary African green monkey kidney cells. Recombinant receptor  
 CC can be produced in large amounts by expression of the nucleic acid  
 CC in prokaryotic or eukaryotic host cells or in transgenic animals.  
 CC It can be used in the detection, prevention and treatment of HAV  
 CC infection, in methods of purifying or removing HAV from samples,  
 CC and methods of determining the anti-HAV activity of cpds.  
 CC Expression in transgenic animals provides a method for testing  
 CC vaccine efficacy.  
 SQ Sequence 451 AA;

Query Match 7.1%; Score 307.5; DB 1; Length 451;

Best Local Similarity 41.6%; Pred. No. 5.9e-10;  
 Matches 87; Conservative 4; Mismatches 53; Indels 65; Gaps 10;

Qy 94 GPPAET-----TOAPATTOAPTTOAPTTOAPTTOAPTTOAPT---TTQAPT 139  
 Db 125 GPPRVTPIVRTVSTVPTTTLPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTT 183  
 Qy 140 QAPTTOAPTTOAPT---TTQAPTTOAPTTOAPT---QAPTTOAPTTOAPTTO 191  
 Db 184 TVPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTT 242  
 Qy 192 APTTITQAATTPAAT-----PAATT-PAATT-PAATT-----PAAT 226  
 Db 243 LPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTTLPPTTTTLP 302  
 Qy 227 TPGVP-----APTSAPVW 239







**THIS PAGE BLANK (USPTO)**

**THIS PAGE BLANK COPY**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 11:36:35 ; Search time 1043.53 Seconds  
(without alignments)  
4123.717 Million cell updates/sec

Title: US-09-294-663-2  
Perfect score: 2821  
Sequence: 1 GAAAGATATACCGACGAC.....TAAAAAAAAAAAAAAAAAAAAA 2821

Scoring table: IDENTITY\_NUC

Searched: 2002476 seqs, 762712212 residues

Database : EST:\*

- 1: em\_est1:\*
- 2: em\_est2:\*
- 3: em\_est3:\*
- 4: em\_est4:\*
- 5: em\_est5:\*
- 6: em\_est6:\*
- 7: em\_est7:\*
- 8: em\_est8:\*
- 9: em\_est9:\*
- 10: gb\_est1:\*
- 11: gb\_est2:\*
- 12: gb\_est3:\*
- 13: gb\_est4:\*
- 14: gb\_est5:\*
- 15: gb\_est6:\*
- 16: gb\_est7:\*
- 17: gb\_est8:\*
- 18: gb\_est9:\*
- 19: gb\_est10:\*
- 20: gb\_est11:\*
- 21: gb\_est12:\*
- 22: gb\_est13:\*
- 23: gb\_est14:\*
- 24: gb\_est15:\*
- 25: gb\_est16:\*
- 26: gb\_est17:\*
- 27: gb\_est18:\*
- 28: gb\_est19:\*
- 29: gb\_est20:\*
- 30: gb\_est21:\*
- 31: gb\_est22:\*
- 32: em\_est11:\*
- 33: em\_est12:\*
- 34: em\_est13:\*
- 35: em\_est14:\*
- 36: em\_est15:\*
- 37: em\_est16:\*
- 38: em\_est17:\*
- 39: em\_est18:\*
- 40: em\_est19:\*
- 41: em\_est20:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	107.6	3.8	1075	14 W29419	W29419 mb98hl1.r1
C 2	101.4	3.6	538	30 A114969	A114969 ui41b06.y

3	101.2	3.6	814	30	A1131654
C 4	100.8	3.6	759	13	W28277
5	90.4	3.2	616	31	AU034820
C 6	90.2	3.2	452	22	AA589477
C 7	90.2	3.2	514	26	AA798859
8	90	3.2	672	31	AU034104
C 9	89.6	3.2	482	24	AA664125
10	89.4	3.2	690	31	AU033402
11	89.4	3.2	566	31	AU033450
12	89.4	3.2	574	31	AU033594
13	89.4	3.2	664	31	AU033617
14	89.4	3.2	603	31	AU033659
15	89.4	3.2	586	31	AU033708
16	89.4	3.2	578	31	AU033712
17	89.4	3.2	672	31	AU033776
18	89.4	3.2	605	31	AU033778
19	89.4	3.2	638	31	AU033787
20	89.4	3.2	622	31	AU034132
21	89.4	3.2	565	31	AU034176
22	89.4	3.2	773	31	AU034671
23	89.4	3.2	630	31	AU034708
24	89.4	3.2	575	31	AU034754
25	89.4	3.2	686	31	AU034976
C 26	88.8	3.1	1222	13	W42209
27	88.4	3.1	636	10	T55873
C 28	88	3.1	355	26	AA798861
29	87.8	3.1	459	23	AI285778
C 30	86.6	3.1	967	14	W34662
C 31	86.2	3.1	422	25	AA673536
32	85.8	3.0	586	31	AU033861
33	85.6	3.0	684	20	C25558
C 34	85	3.0	408	21	AA484400
35	84.4	3.0	555	31	AU033469
36	84.4	3.0	521	31	AU033607
37	84.2	3.0	558	31	AU033502
C 38	83.8	3.0	1461	14	W15723
39	83.8	3.0	576	26	AA754683
C 40	83.6	3.0	1257	13	W98479
C 41	83.6	3.0	1311	14	W15735
C 42	82.8	2.9	1161	13	W41959
C 43	82.4	2.9	558	31	AU034078
44	82	2.9	714	20	AA519116
45	81.8	2.9	498	31	AI253548

ALIGNMENTS

RESULT 1

W29419/c

LOCUS

DEFINITION

ACCESSION

NID

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

W29419 mb98hl1.r1 Soares mouse p3NMF19.5 Mus musculus cDNA clone 337509  
5', mRNA sequence.

W29419  
GI309566

EST.  
house mouse.

Mus musculus  
Eukaryotae; mitochondrial eukaryotes; Metazoa; Chordata; Vertebrata; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

1 (bases 1 to 1075)  
Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenger, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.

The WashU-HHMI Mouse EST Project  
Unpublished (1996)

Contact: Marra M/Mouse EST Project  
WashU-HHMI Mouse EST Project  
Washington University School of Medicine

11-SEP-1996



Query Match	3.68	Score 101.4	DB 30	Length 538
Best Local Similarity	55.1%	Pred. No. 9.1e-11		
Matches 220	Conservative 0	Mismatches 176	Indels 3	Gaps 1
QY 1629	CTCCGGCCCCACCGCTGCGCCCGACCGCAGCTCTCTACTACTGCTGCGCCCTGAATCCCAA	1688		
DB 525	CTGTTGCTGCCACCTGCTGCTCCACCACTGCTGCCCCCACTGCTACACAGCAGCTGCTG	466		
QY 1689	CCAAGTGTACAGTACCACTACTGCTGCTCCACCGCAGCCCTACTACTGCTGTCCCTGT	1748		
DB 465	CCACCGCTACTGCCACCGCTACTGCCACCACTGCTGCTGCCACAGCACTGCTGCCACCG	406		
QY 1749	AAATCCCAATCACTGTGCATCAGGGCTACCGGTGCCCGCCACCGCTGCCCGCCACCGGT	1808		
DB 405	CTGCTGTGCCACAGCACTGCTGCCACCGCTGCTGCCACAGCACTGCTGCCACCGCTG	346		
QY 1809	CCCCACCGCAGCCCTACTACTGCTGCCAGAAATCCCAACTACTGTACATCAACCAC	1868		
DB 345	CTGCCACCGCTGCTGCCACCGCTGCTGCCACAGCACTGCTGCCACCGCTGCTGCCACAG	286		
QY 1869	CTACTGTGCCCCCACTACCGCAGCACCTGCTGCCCAACACCACTACTGTACCAACCA	1928		
DB 285	CAACTGTGCCACCGCTCTACAGCATCCAGAGCTGTGACGTGGCGACGAGAGA---G	229		
QY 1929	CTGTGCCCCCTACTACCGCAGCACCTGCTGCCAACACACACAGTACTGACACCAACCG	1988		
DB 228	AATGGGGCTTGTGGTGGCTCAGACAGCGCCACACATCCAGAGCTGCAGCAGCCCCCG	169		
QY 1989	CAGCCCTACTACCGCAGCACCTGCTGCCCAACACCAACAG	2027		
DB 168	GAGCTGGAACACAGCAGCCCCAGAACCCAGGCTACAG	130		

  

LOCUS	A1131654	814 bp	mRNA	EST	14-SEP-1998
DEFINITION	u141806.x1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884947				
	3' similar to TR:014633 014633 SKIN-SPECIFIC PROTEIN. ; mRNA				
ACCESSION	A1131654				
NID	g3601670				
KEYWORDS	EST.				
SOURCE	house mouse.				
ORGANISM	Mus musculus				
REFERENCE	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;				
AUTHORS	Rodentia; Sciurognathi; Muridae; Murinae; Mus.				
	1 (bases 1 to 814)				
	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T.,				
	Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M.,				
	Schellensberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B.,				
	Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and				
	Waterston, R.				
TITLE	The WashU-HHMI Mouse EST Project				
JOURNAL	Unpublished (1996)				

Contact: Marra M/Mouse EST Project  
WASHU-HHMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium (info@image.llnl.gov) for further information.  
MGI:969271  
Possible reversed clone: similarity on wrong strand  
Seq primer: custom primer used  
High quality sequence stop: 375.

```

FEATURES
  source      .
  Location/Qualifiers
    1. .814
      /organism="Mus musculus"
      /strain="C57BL"
      /note="Vector: pMEL18S-FL3; Site_1: DraIII (CACTGTCGTG);

```

```

Site_2: DraIII (CACCAATG); 1st strand cDNA was primed
with an oligo(dT) primer [ATGTGGCCTTTTTTTTTTTTTTTT];
double-stranded cDNA was ligated to a DraIII adaptor
[TCGTGGCCTACAGG], digested and cloned into distinct DraIII
sites of the pME18S-FL3 vector (5' site CACTGTGTG 3' site
CACCAATG). XhoI should be used to isolate the cDNA
insert. Size selection was performed to exclude fragments
<1.5kb. Library constructed by Dr. Sumio Sugano
(University of Tokyo Institute of Medical Science).
Custom primers for sequencing: 5' end primer
CTTCTGCTCTAAAAGCTGCG and 3' end primer
CGACCTGCAGCTCGAGCAC.
/db_xref="taxon:10090"
/clone="1884947"
/clone_lib="Sugano mouse embryo mewa"
/dev_stage="embryo, 14 dpc"
/lab_host="DH10B"
184 a 245 c 211 g 173 t 1 others

```

Query Match	3.58;	Score	101.2;	DB	30;	Length	814;
Best Local Similarity	55.08;	Pred.	No. 9.8e-11;				
Matches	221;	Conservative	0;	Mismatches	178;	Indels	3;
Gaps	1;						
QY	1629	CTCGGGCCCCACCGCTGCCCCACCGCAGTCTCTACTACTGCTGCCCTCGAATCCCCAA	1688				
Db	337	CTGTGTGCGCAGTGTCTGTGCCACCACTGCTGCCCCACTGCTTACCACAGCAGCTGCTG	396				
QY	1689	CCACTGTACAGTATACCACCTACTGCTCTCCACCGCAGCCCTTACTACTGCTGTCCCTG	1748				
Db	397	CCACCGCTACTGCCACCGTTACTGCCACCACTGCTGTGCCACAGCAACTGCTGCCACCG	456				
QY	1749	AAATCCCAATCATGTCACATCAGCGCTACCGTGCCTGCCCCACCGCTGCCCCACCGCTG	1808				
Db	457	CTGCTGTGCCACAGCAACTGCTGCCACCGTGTGCTGCCACAGCAACTGCTGCCACCGTG	516				
QY	1809	CCCCACCGCAGCCCCCTACTACTGCTGTCCAGAAATCCCAACTACTGTACATCATCCAC	1868				
Db	517	CTGCCACCGTGTGTGCCACCGCTGCTGCCACAGCAACTGCTGCCACCGCTGCTGCCACAG	576				
QY	1869	CTACTGTGTCCTCCACTTACCGCAGCAGCTGCCCCACACACACAGTACTGTATCCACCCA	1928				
Db	577	CAACTGTGCCACCGCTGTACAGCATCCAGAGCTGTGACGTTCGGCCACGGAGAGA---G	633				
QY	1929	CTGTGTGCCCTTACTACCGCAGCAGCTGCTCCCAACACCAAGTGTACTGTCCACCCACCG	1988				
Db	634	AATCGGGCTGTGTGTGCTCAGACAGCGCCACCATCCAGAGCTGCAGCAGCCCCCG	693				
QY	1989	CAGCCCCCTACTACCGCAGCAGCTGCCCCCAACACCAACCATCA	2030				
Db	694	GTGGTGAACCCACAGCAGCCCCCAGAACCCAGAGCTCAGGCTACAGCCA	735				

RESULT	4				
W28277/c					
LOCUS					
DEFINITION					
W28277					
CDNA, mRNA sequence.					
W28277					
g1308243					
EST.					
human.					
Homo sapiens					
Eukaryotae; mitochondrion					
Eukaryotae; Metazoa; Chordata;					
Vertebrata; Eutheria; Primates;					
Catarrhini; Hominoidea; Homo.					
1 (bases 1 to 759)					
Macke, J., Smallwood, P. and Nathans, J.					
Adult Human Retina cDNA					
Unpublished (1996)					
CONTACT					
Dr. Jeremy Nathans					
Dr. Jeremy Nathans, Dept. of					
Molecular Biology and Genetics					

Johns Hopkins School of Medicine  
725 North Wolfe Street, Baltimore, MD 21205  
Tel: 410 955 4678  
Fax: 410 614 0827  
Email: jeremy.nathans@gmail.bs.jhu.edu  
Clones from this library are NOT available.  
PCR Primers  
FORWARD: CTTTGGACAGTTCAGCCGTGGTAAAT  
BACKWARD: GAGTGCGCTATGACATATTCCTCAGGGTAA  
Seq primer: GGCTAAAAGACAAAGAAAT.

source

```

1. .759
/organism="Homo sapiens"
/notes="Organ: eye; Vector: lambda gt10; Site_1: ECORI;
Site_2: EcoRI; The library used for sequencing was a
sublibrary derived from a human retina cDNA library.
Inserts from retina cDNA library DNA were isolated,
randomly primed, PCR amplified, size selected, and cloned
into lambda gt10. Individual plaques were arrayed and
used as templates for PCR amplification, and these PCR
products were used for sequencing."
/db_xref="taxon:9606"
/clone_lib="Human retina cDNA randomly primed sublibrary"
/sex="mixed (males and females)"
/tissue_type="retina"
/dev_stage="adult"
/lab_host="E. coli strain K802"
<1. .759
38 a 71 c 494 g 16 t 140 others

```

Query Match	3.6%	Score 100.8;	DB 13;	Length 759;
Best Local Similarity	41.8%;	Pred. No. 1.2e-10;		
Matches 272;	Conservative 0;	Mismatches 377;	Indels 1;	Gaps 1;
QY 1542	CCCAACTGCGGACACCTGCACCGTGTCCACCTGCACAGCCACCGCGAACTGCAGGCC	1601		
DB				
QY 732	CCCCCGCNCNCCCGCCCGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCG	673		
DB				
QY 1602	CAGTCCCACTACAACGCAATTCTACTCGGCGCCCAACCGCTGCCCGCCACCGCAGCTC	1661		
DB				
QY 1662	CTACTACTGTGCCCCGTGAATCCCAACCACTGTCAAGTACCAGCTACTGCTGTCCCA	1721		
DB				
QY 612	GCNCGCNCGGCGCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCC	553		
DB				
QY 1722	CCGAGCGCCCTACTACTGCTGTGCTGAAATCCCAATCCACTGTGCATCAGCGCTACCG	1781		
DB				
QY 552	GCCCCGCGCNCGCCGCGC-NCCGCGCNCGCCGCGCGCGCGCGCGCGCGCGCGCG	494		
DB				
QY 1782	CTGCCCCACCGTGTGCCCCCAACCGCTGCCCGCCAGCGACGCCCTACTACTGTCTCCAG	1841		
DB				
QY 493	GCGCCCGCCCGCGCNCGCCGCGCNCGCCGCGCGCGCGCGCGCGCGCGCGCGCG	434		
DB				
QY 1842	AAATCCCACTACTGTGCATCACACGCTACTGTGCGCCCACTACCGCAGCACTGTGCC	1901		
DB				
QY 433	CNCGNCCCGCCCGCCCGCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG	374		
DB				
QY 1902	CCAAACACCACTACTGTACACCCACTGTGTGCCCTACTACGCGACGACCTGTCTCCA	1961		
DB				
QY 373	CCNCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG	314		
DB				
QY 1962	ACACCACTGTACTGTACACCGCGCGCCCTACTACCGCAGCACTGCCCGCCACA	2021		
DB				
QY 313	CCGCNCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCGCCCG	254		
DB				
QY 2022	CCACAGTCACTGTACCACTGCTGCCCGCACTGCAGCTGCCCGCTACCGTCCGACCTG	2081		
DB				
QY 253	ANNAGNCTCCGNCNCGNNGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT	194		
DB				
QY 2082	CACCCACACCACTGCGCCCGGTAACTACAACGACGCGCACAGCTACCACTGAAG	2141		
DB				

[illegible]

BASE COUNT	185 a	194 c	181 t	1 others
ORIGIN	185 a	194 c	181 t	1 others

Query Match	3.2%;	Score 90.4;	DB 31;	Length 616;
Best Local Similarity	50.3%;	Pred. No. 1.3e-08;		
Matches: 220; Conservative		0; Mismatches 217;	Indels	Gaps

QY	1482	CATCGACTGTCGCCCGACAGCTGTCTCCAAACGGCTGCCCCAACGGCTGCCCCAAACGGGTG	1541
Db	44	CTTCTGCTGTTTCATCTTCCACATCATCATCTGCTGCAAGCTCTTCTGCATCTTCATCTG	103
QY	1542	CCCCAACTGCGGACGCTCCACCGTGGTCCCACTGCAAGCGCACCGCAACTGCAGGCC	1601
Db	104	CTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCTTCCCATCTTCATCTGCATCAT	163
QY	1602	CAGTCCCACTATCAACCGCAATTCTACTTCGGGCCCCAACCGCTGCCCCACCCAGCTC	1661
Db	164	CAAGCTCTTCCCATCTTTCATCTGCATCATCAAGCTCTGCCCATCTTCTCATCTGCATCAT	223
QY	1662	CTACTACTGTCCTCCCTGAATCCCCAACCACTGTCACAGTACCACCTACTGCTGCTCCCA	1721
Db	224	CAAGCTCTGCTCATCTTCATCTGCATCATCTTCATCTGCATCATCTTCATCTGCATCAT	283
QY	1722	CCGACGCCCTCTACTGCTGTCCCTGAAATCCCAATCCCACTGTACACATCAGCGGCTACCG	1781
Db	284	CTGCACAACTACAGCTGCAACTACAATTTGCAACTACAGCTGCAACTACAACCTGCAACTA	343
QY	1782	CTGCCCGCACCGCTGCCCGCCAGCGTGGCCCCACCGAGCGCCCTACTACTGCTGTCCCGAG	1841
Db	344	CAACCGCAACCCAGCTACCAACCAAGCGCAACCAACCAAGCTACCAACCAAGCTGCAACCA	403
QY	1842	AAATCCCAACTACTGTTCATCACTCACCACCTACTGCTGCCGCCCACTACCGCAGCACTGCC	1901

Db 404 TAGNTACCAACCAGCTGGCAACTGCAACTGCAACTATACAACCGCAACCAACGCTACAACCA 463

QY 1902 CCAACACCAACGTCAGTCACT 1918  
+ + + + +  
Db 464 CAGCTACAACCAACACT 480

RESULT 6  
AA589477/c 452 bp mRNA EST 16-SEP-1997  
LOCUS v147hl2.s1 Stratagene mouse skin (#937313) Mus musculus CDNA clone  
DEFINITION 975431 3', mRNA sequence.  
ACCESSION AA589477  
NID 92402857  
KEYWORDS EST.  
SOURCE house musculus  
ORGANISM Mus musculus  
Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 452)  
AUTHORS Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.  
TITLE The WashU-HMI Mouse EST Project  
JOURNAL Unpublished (1996)  
COMMENT Contact: Marra M/Mouse EST Project  
WashU-HMI Mouse EST Project  
Washington University School of MedicineP  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: mouseest@watson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
MGI:556159  
Possible reversed clone: polyT not found  
Seq primer: -28ml3 rev1 ET from Amersham  
High quality sequence stop: 447.  
Location/Qualifiers  
.452  
/organism="Mus musculus"  
/strain="C57BL/6"  
/note="Organ: skin; Vector: pBluescript SK-; Site:1:  
ECORI; Site:2: XhoI; Cloned unidirectionally. Primer:  
Oligo df. Whole skin from 11 week old C57BL/6 female mice.  
Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5'  
adaptor sequence: 5' GAATTCGGCACGAG 3' -3' adaptor  
sequence: 5' CTCGAGTTTTTTTTTTTTTTTTT 3'"  
/db\_xref="taxon:10090"  
/clone="975431"  
/clone\_lib="Stratagene mouse skin (#937313)"  
/sex="females"  
/tissue\_type="whole skin"  
/dev\_stage="11 weeks old"  
/lab\_host="SOLR (kanamycin resistant)"  
complement(<1..>452)

mRNA 32 a 93 c 193 g 134 t  
BASE COUNT  
ORIGIN

Query Match 3.2% Score 90.2; DB 22; Length 452;  
Best Local Similarity 54.3%; Pred. No. 1.4e-08;  
Matches 187; Conservative 0; Mismatches 153; Indels 0; Gaps 0;

QY 1503 CTGTCTCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCA 1562  
+ + + + +  
Db 346 CTTTGTGTGAACAGCCCTTCCCACAGCCACGCGACGCGACGACGACGACGACGACGACGACG 287  
+ + + + +  
QY 1563 CGGTGGTGCCACCTGTGCACACGGCCACCGGCACTGTGACGCCCGCCAGTCCCACTTACAACCGCAA 1622  
+ + + + +



[illegible]

RESULT	8
AU034104	
LOCUS	
DEFINITION	672 bp mRNA
ACCESSION	D19441
NID	Q3799528
KEYWORDS	EST.
SOURCE	Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA, clone_lib:SL.
ORGANISM	Dictyostelium discoideum
REFERENCE	Eukaryota; Dictyostellida; Dictyostellum.
AUTHORS	1 (bases 1 to 672)
TITLE	Urushihara,H
JOURNAL	Developmental cDNA in Dictyostelium discoideum
REFERENCE	Published Only in database (1998) In press
AUTHORS	2 (bases 1 to 672)
TITLE	Urushihara,H.
JOURNAL	Direct Submission
	Submitted (23-SEP-1998) to the DDBJ/EMBL/GenBank databases..HidE
	Urushihara, University of Tsukuba, Institute Of Biological
	Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
	(E-mail:qxurushi@bank.dna.afrc.go.jp, Tel:+81-0298-53-4664,
	Fax:+81-0298-53-6614)
COMMENT	PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES	Location/Qualifiers
	1..672
	source

	Location/Vouchers	
	i. .672	
	/organism="Dictyostelium discoideum"	
	/strain="AX4"	
	/db_xref="taxon:44689"	
	/clone_lib="SL"	
	/dev_stage="S1"	
	214 a    207 c    60 g    190 t    1 others	
	BASE COUNT	
	ORIGIN	
	source	
	FEATURES	

	Query Match	Score 90;	DB 31;	Length 672;
	Best Local Similarity	50.1%;	Pred. No. 1.6e-08;	
	Matches 222;	Conservative	0;	Mismatches 221;
				Indels 0;
				Gaps 0;
QY	1482	CATCGACTGCTGCCCGGACAGCTCTCCACGGCTGCCCAACGGCTGCCCAACGGCTG	1541	
Db	78	CATCTTCACCATCATCATCTGCTGCTCAAGCTCTTCTCCATCTTCATCTGCTGCTCAAGCTCTT	137	
QY	1542	CCCAACTGCGGACACCTCTCCACCGTGTGCCACCTGCGCAACGCCACCGCACTCGAGGCC	1601	

Db	138	CTCCATCTTCATCTGCTGCAGCTCTTCCGCATCTTTCATCTGCATCATCAAGCTCTTCCC	197
Qy	1602	CAGTCCACACCTACAACCGCAATTCCTACTCGGGCCCCACCGGTGCCCCACCCAGCAGCT	1661
Db	198	CATCTTCATCTGCATCATCAAGCTCTGCCCATCTTTCATCTGCATCATCAAGCTCTGCTC	257
Qy	1662	CTACTACTGCTGCCCCGGAATCCCAACCACTGTCAAGACTACCACTACTGCTGCTGCCA	1721
Db	258	CATCTTCATCTGCATCATCTTTCATCTGCATCATCTTTCATCTGCATCATCTGCGAACCTA	317
Qy	1722	CCGAGGCCCCCTACTACTGCTGTGCCCTGAATCCCAACTCTGTACACATCAGCGCTACCG	1781
Db	318	CAGCTGCACTACAAATGGCAACTACAGCTGCAACTCAACTGCNACTACAAACCCGCACTA	377
Qy	1782	CTGCCCCCAGCGGTGCCCCACCGCTGCCGCCACCGCGGCCCTTACTGCTGTCCGAG	1841
Db	378	CAGCTACAAGCCACGAGCCACACCACTACACCACTACACCACTGCAACCATAGCTACAACCA	437
Qy	1842	AAATCCCAACTACTGTGCATATCACACTACTGCTGCCCCACTACTACCGGACGACCTGCC	1901
Db	438	CAGCTGCACTACAACTGCACTACAAACCGGACCACTACAAACCCAGCTACAACCA	497
Qy	1902	CCAAACCAACGTCACGTGTACCA	1924
Db	498	CANCTACCTCTTAAATTTAATCA	520

RESULT	9
AA664125/c	
LOCUS	12-NOV-1997
DEFINITION	EST
	mRNA
	482 bp
	AA664125
	ac06c04.s1 Stratiogene lung (#937210) Homo sapiens CDNA clone 856564
	3', similar to contains element MSR1 MSR1 repetitive element ; , mRNA sequence.
ACCESION	AA664125
NID	G2618116
KEYWORDS	EST.
SOURCE	human.
ORGANISM	Homo sapiens
	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
	Primates; Catarrhini; Homnidae; Homo.
REFERENCE	1 (bases 1 to 482)
AUTHORS	Hillier,V., Allen,M., Bowles,L., Dubucque,T., Geisel,G., Jost,S., Krizman,D., Kucaba,T., Lacy,M., Le,N., Lennon,G., Marra,M., Martin,J., Moore,B., Schellenberg,K., Steptoe,M., Tan,F., Theising,B., White,Y., Wylie,T., Waterston,R. and Wilson,R.
TITLE	WashU-NCI human EST Project
JOURNAL	Unpublished (1997)

Contact: Wilson RK  
Washington University School of Medicine  
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108  
Tel: 314 286 1800  
Fax: 314 286 1810  
Email: estewartson.wustl.edu  
This clone is available royalty-free through LLNL ; contact the  
IMAGE Consortium ([info@image.llnl.gov](mailto:info@image.llnl.gov)) for further information.  
Seq primer: -40ml3 fwd. ET from Amersham  
High quality sequence stop: 425.

```

FEATURES
source
align quality sequence stop: 425.
Location/Qualifiers
1. .482
/organism="Homo sapiens"
/note="Organ: lung; Vector: pBluescript SK-; Site: 1:
EcoRI; Site: 2: XhoI; Cloned unidirectionally. Primer:
Oligo dt. normal lung. Average insert size: 1.0 kb;
Uni-ZAP XR Vector: -5' adaptor sequence: 5' GAATTCGGCAGCAG
3' -3' adaptor sequence: 5' CTCGAGTTTTTTTTTTTTTTT 3'"
/db_xref="taxon:9606"
/clone="855854"
/clone.lib="Stratagene lung (#937210)"
/sex="male"
/dev_stage="72 years"
/lab_host="SOLR cells (kanamycin resistant)"

```

```
BASE COUNT      51 a   113 c   176 g   142 t
ORIGIN

Query Match      3.2%; Score 89.6; DB 24; Length 482;
Best Local Similarity 50.0%; Pred. No. 1.9e-08;
Matches 224; Conservative 0; Mismatches 224; Indels 0; Gaps 0;

QY 1485 CGACTGCTGCCCGAGAGCTGCTCCAAAGGCTGCCCGAAGGCTGCCCGAAGGCTGCCCG 1544
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 459 CTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 400
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1545 CAACCTCGCGCACCTCCACCGTGGTCCCACTGCAAGCGCACCGCAACCTGCAAGCGCCAG 1604
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 399 CAGCAGCAGTTCCACAGAGTAGTGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 340
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1605 TCCACCTCAACACCGCAATTCCTTACTCCGGCCCCCACCCTGCGCCCGCAGCTCCCTA 1664
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 339 CAGCAGCATCTAATTAATTCATCATCAATCAAAATCAGCAACAGATACAGCAGCAGCAACAG 280
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1665 CTACTGCTGCCCTGAATCCCAACCACTGTACAGTACCACTACTGTGCTGCCACCG 1724
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 279 CAGCTGCGGGAATAGCAGAGCTGCGAGCTCCAAACAACAGCAAGCAGCAGCAGCAGCAGCAG 220
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1725 CAGCCCTACTACTGCTGCTCCCTGAATCCCACTCACTGTCAATCAGCGCCTACCGCTG 1784
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 219 CAGCAGCAGCAGCAGGAGGCTTTCAGGCGCCGACCAATTCAGCAGCAGCAGCAGCAGCAG 160
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1785 CCCCCACCGTGCCTCCACCGCTGCCCGCCACCGCAGCGCCCTACTGTGCTCCAGAAA 1844
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159 CAGCAGCAGCTCCGCTCCAGGCTCTGCCCGCAGCAGCTGCGCAGCAGTGCATCACACA 100
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1845 TCCCACTACTGTACATCACTACCACTACTGCTGCCCGCCACTACCGCAGCAGCTGCCCGCA 1904
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 99 CAGCAGCAGCAGCGCCGACCAACAGCGCCAGCAGCAGCTCCAGTTGCTCAGAACCAACCATCA 40
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1905 ACACCAGTCATCTACCAACCACTGTC 1932
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 39 CAATCCCGCCAGCAGTCGCGAGCCCGAGC 12
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 10
AU033402
LOCUS      AU033402      690 bp      mRNA      EST      28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLA726, mRNA sequence.
ACCESSION  AU033402
NID        g3798826
KEYWORDS   EST.
SOURCE     Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
           clone_lib:SL.
ORGANISM   Dictyostelium discoideum
REFERENCE   1. (bases 1 to 690)
            Eukaryota; Dictyosteliida; Dictyostelium.
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2. (bases 1 to 690)
            Urushihara,H.
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel: +81-0298-53-4664,
            Fax: +81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
            Location/Qualifiers
            source
            1..690
               /organism="Dictyostelium discoideum"
               /strain="AX4"
               /db_xref="taxon:44689"
               /clone_lib="SL"
               /dev_stage="slug"
BASE COUNT      212 a   213 c   67 g   198 t
ORIGIN

Query Match      3.2%; Score 89.4; DB 31; Length 690;
Best Local Similarity 50.1%; Pred. No. 2.1e-08;
Matches 222; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

QY 1482 CATCGACTGCTGCCCGAGAGCTGCTCCAAAGGCTGCCCGAAGGCTGCCCGAAGGCTG 1541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 100 CATCTTCACATCATCATCTGCTGCAAGCTCTTCTCCATCTTCATCTGCTGCAAGCTCT 159
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1542 CCCCRAACTCGGCACCTCCACGGTGGTCCCACTGCAAGCGCACCGCAACTGCAAGCCC 1601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 160 CTGCATCTTCATCTGCTGCAAGCTTTCCTCCATCTTCATCTGTCATCATCAAGCTCTTCCC 219
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1602 CAGTCCCACTCAACCGCAATTCCTTACTCCGGCCCCCACCCTGCGCCCGCAGCTC 1661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 220 CATCTTCATCTGCATCATCAAGCTCTGCCCATCTTTCATCTGTCATCATCAAGCTCTGCTC 279
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1662 CTACTACTGCTGCCCTGAATCCCAACCACTGTACAGTACCACTACTGTGCTGCCCA 1721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 280 CATCTTCATCTGCATCATCTTTCATCTGTCATCATCTTTCATCTGTCATCATGCAAGCACTA 339
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1722 CCGCAGCCCCCTACTACTGCTGCTCCCTGAATCCCACTACTGTCACTCAGCGCCTACCG 1781
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 340 CAGCTGCAACTACAATGCAACTACAGCTGCAACTACAACCTGCAACTACAACCGCAACCA 399
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1782 CTGCCCCCAGCGTGCCTCCAGCGTGCCTGCCCGCCACCGCAGCGCCCTACTGTGCTCCAG 1841
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 400 CAGCTACACACAGCCACCAACCAACAGCTACACACAGCTGCAACCATAGTACAAACA 459
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1842 AAATCCCACTACTGTGCATCACCACCTACTGTGCTGCCCGCCACTACCGCAGCAGCTGCC 1901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 460 CAGCTGCAACTACAACCTGCAACTACAACCGCAACCAACAGCTACAAACCAACAGCTACAACA 519
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 1902 CCAACACACAGTCAGTGTACCA 1924
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 520 CAGCTACCTCTTAAATTTAATCA 542
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 11
AU033450
LOCUS      AU033450      566 bp      mRNA      EST      28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLA841, mRNA sequence.
ACCESSION  AU033450
NID        g3798874
KEYWORDS   EST.
SOURCE     Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
           clone_lib:SL.
ORGANISM   Dictyostelium discoideum
REFERENCE   1. (bases 1 to 566)
            Eukaryota; Dictyosteliida; Dictyostelium.
AUTHORS     Urushihara,H.
TITLE       Developmental cDNA in Dictyostelium discoideum
JOURNAL     Published Only in DataBase (1998) In press
REFERENCE   2. (bases 1 to 566)
            Urushihara,H.
AUTHORS     Urushihara,H.
TITLE       Direct Submission
JOURNAL     Submitted (22-SEP-1998) to the DDBJ/EMBL/GenBank databases. Hideko
            Urushihara, University of Tsukuba, Institute of Biological
            Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
            (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel: +81-0298-53-4664,
            Fax: +81-0298-53-6614)
COMMENT     PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
            Location/Qualifiers
            source
            1..566
               /organism="Dictyostelium discoideum"
               /strain="AX4"
               /db_xref="taxon:44689"
               /clone_lib="SL"
               /dev_stage="slug"
BASE COUNT      182 a   179 c   46 g   159 t
ORIGIN
```

Query Match		3.2%;	Score 89.4;	DB 31;	Length 574;
Best Local Similarity		50.1%;	Pred. No. 2.1e-08;		
Matches 222;		Conservative	0;	Mismatches 221;	Indels 0; Gaps 0;
QY	1482	CATCGACTGTCGCCCGACAGCTGCTCCAAAGCGTGC	CCCAACGGCTGCCCAACGGCTGCCCAACGGCTG	1541	
Db	14	CATCTTCACCATCATCATCTGCTGCAAGCTCTTCT	CCATCTTTCATCTGCTGCAAGCTCTT	73	
QY	1542	CCCAACTGCGGACCCCTCCACCGTGTCCACCTGCA	AGCCGACCCGCAACTGCGAGCCG	1601	
Db	74	CTCCATCTTTCATCTGCTGCAAGCTCTTCCCATCT	TTCATCTGCAATCATCAAGCTCTTCCC	133	
QY	1602	CAGTCCACCATACAAACCGCAATTCCTACTCCGCG	CCCCCAGCGTGC	1661	
Db	134	CATCTTCATCTGCATCATCAAGCTCTGCCCATCT	TTCATCTGCATCATCAAGCTCTGCTC	193	
QY	1662	CTACTACTGTCGCCCTGAAATCCCAACCACTGTCA	CAGTACCACTGCTGCTGCCCA	1721	
Db	194	CATCTTCATCTGCATCATCTTTCATCTGCATCAT	CTTTCATCTGCATCATCTGCAAGCTA	253	
QY	1722	CCGAGCGCCCTACTACTGCTGTCCCTGAAATCCCA	ATCCCAATCAGTGTACATCAGCGCTACCG	1781	
Db	254	CAGCTGCAACTACAATTTGCAACTACAGCTGCAAC	TGCAACTACAACTGCAACTACAAACCGCAACCA	313	
QY	1782	CTGCCCCACCGTGC	CCCCACCGTGC	1841	
Db	314	CAGCTAACACCAACAGCCACCAACAGCTACAA	CCAGCTACAAACAGCTACAAACAGCTACAAACCA	373	
QY	1842	AAATCCCAACTACTGTCAATCACCACCTACTGCT	GCCGCCCTACTGCTGCCCGCACTACCGACGCACTG	1901	
Db	374	CAGCTGCAACTACAACCTGCAACTACAAACCG	CAACCGCAACAGCTACAAACAGCTACAAACCA	433	
QY	1902	CCAAACACCACTGCTGTACCA	1924		
Db	434	CAGCTACCTCTTAAATTTAATCA	456		
RESULT 13					
LOCUS	AU033617	664 bp	mrna	EST	28-OCT-1998
DEFINITION	Dictyostellium discoideum slug cDNA, clone SLB234, mRNA sequence.				
ACCESSION	AU033617				
NID	g3799041				
KEYWORDS	EST.				
SOURCE	Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA, clone_lib:SL.				
ORGANISM	Dictyostellium discoideum				
REFERENCE	1 (bases 1 to 664)				
AUTHORS	Urushihara,H.				
TITLE	Developmental cDNA in Dictyostellium discoideum				
JOURNAL	Published Only in Database (1998) In press				
REFERENCE	2 (bases 1 to 664)				
AUTHORS	Urushihara,H.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-SEP-1998) to the DBJ/EMBL/GenBank databases. Hideko Urushihara, University of Tsukuba, Institute of Biological Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664, Fax:+81-0298-53-6614)				
COMMENT	PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.				
FEATURES	Location/Qualifiers				
source	1..664				
	/organism="Dictyostellium discoideum"				
	/strain="AX4"				
	/db_xref="taxon:44689"				
	/clone_lib="SL"				
	/dev_stage="slug"				
BASE COUNT	193 a	213 c	66 g	192 t	
ORIGIN					

Query Match		3.2%;	Score 89.4;	DB 31;	Length 566;
Best Local Similarity		50.1%;	Pred. No. 2.1e-08;		
Matches 222;		Conservative	0;	Mismatches 221;	Indels 0; Gaps 0;
QY	1482	CATCGACTGTCGCCCGACAGCTGCTCCAAAGCGTGC	CCCAACGGCTGCCCAACGGCTGCCCAACGGCTG	1541	
Db	6	CATCTTCACCATCATCATCTGCTGCAAGCTCTTCT	CCATCTTTCATCTGCTGCAAGCTCTT	65	
QY	1542	CCCAACTGCGGACCCCTCCACCGTGTCCACCTGCA	AGCCGACCCGCAACTGCGAGCCG	1601	
Db	66	CTCCATCTTTCATCTGCTGCAAGCTCTTCCCATCT	TTCATCTGCAATCATCAAGCTCTTCCC	125	
QY	1602	CAGTCCACCATACAAACCGCAATTCCTACTCCGCG	CCCCCAGCGTGC	1661	
Db	126	CATCTTCATCTGCATCATCAAGCTCTGCCCATCT	TTCATCTGCATCATCAAGCTCTGCTC	185	
QY	1662	CTACTACTGTCGCCCTGAAATCCCAACCACTGTCA	CAGTACCACTGCTGCTGCCCA	1721	
Db	186	CATCTTCATCTGCATCATCTTTCATCTGCATCAT	CTTTCATCTGCATCATCTGCAAGCTA	245	
QY	1722	CCGAGCGCCCTACTACTGCTGTCCCTGAAATCCCA	ATCCCAATCAGTGTACATCAGCGCTACCG	1781	
Db	246	CAGCTGCAACTACAATTTGCAACTACAGCTGCAAC	TGCAACTACAACTGCAACTACAAACCGCAACCA	305	
QY	1782	CTGCCCCACCGTGC	CCCCACCGTGC	1841	
Db	306	CAGCTAACACCAACAGCCACCAACAGCTACAA	CCAGCTACAAACAGCTACAAACAGCTACAAACCA	365	
QY	1842	AAATCCCAACTACTGTCAATCACCACCTACTGCT	GCCGCCCTACTGCTGCCCGCACTACCGACGCACTG	1901	
Db	366	CAGCTGCAACTACAACCTGCAACTACAAACCG	CAACCGCAACAGCTACAAACAGCTACAAACCA	425	
QY	1902	CCAAACACCACTGCTGTACCA	1924		
Db	426	CAGTACCTCTTAAATTTAATCA	448		
RESULT 12					
LOCUS	AU033594	574 bp	mrna	EST	28-OCT-1998
DEFINITION	Dictyostellium discoideum slug cDNA, clone SLB210, mRNA sequence.				
ACCESSION	AU033594				
NID	g3799018				
KEYWORDS	EST.				
SOURCE	Dictyostellium discoideum (strain:AX4) slug cDNA to mRNA, clone_lib:SL.				
ORGANISM	Dictyostellium discoideum				
REFERENCE	1 (bases 1 to 574)				
AUTHORS	Urushihara,H.				
TITLE	Developmental cDNA in Dictyostellium discoideum				
JOURNAL	Published Only in DataBase (1998) In press				
REFERENCE	2 (bases 1 to 574)				
AUTHORS	Urushihara,H.				
TITLE	Direct Submission				
JOURNAL	Submitted (22-SEP-1998) to the DBJ/EMBL/GenBank databases. Hideko Urushihara, University of Tsukuba, Institute of Biological Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan (E-mail:qxurushi@bank.dna.affrc.go.jp, Tel:+81-0298-53-4664, Fax:+81-0298-53-6614)				
COMMENT	PROJECT = 'Dictyostellium discoideum cDNA project in Japan'.				
FEATURES	Location/Qualifiers				
source	1..574				
	/organism="Dictyostellium discoideum"				
	/strain="AX4"				
	/db_xref="taxon:44689"				
	/clone_lib="SL"				
	/dev_stage="slug"				
BASE COUNT	182 a	182 c	47 g	162 t	1 others
ORIGIN					

```
Query Match          3.2%; Score 89.4; DB 31; Length 664;
Best Local Similarity 50.1%; Pred. No. 2.1e-08;
Matches 222; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

Qy 1482 CATCGACTGCTGCCCGGACAGCTGCTCCAAAGGGTGCCTCCCAACGGCTGCCCAACGGCTG 1541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 CATCTTCAACATCATCATCTGCTGCAAGCTCTTCTCCATCTTCCATCTGCTGCAAGCTCTT 162

Qy 1542 CCCCAACTGCCGACCTTCACCGTGTCCACCTGCAACGGCCAGCCGCAACTGCGAGCCC 1601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 CTCATCTTTCATCTGCTGCAAGCTCTTCCCATCTTCCCATCTTCCATCTGCAAGCTCTTCCC 162

Qy 1602 CAGTCCCACTTACCAACCGCAATCTCTACCGGCGCCACCGCTGCCCGCCAGCGCTC 1661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 CATCTTCACTGCAATCAAGCTCTGCCCATCTTCCATCTGCAATCAAGCTCTGCTC 282

Qy 1662 CTACTACTGCTGCCCGCTGAATCCCAACCACTGTGCAGTACCACTACTGCTGCCCA 1721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 CATCTTCACTGCAATCACTTTCATCTGCAATCACTTTCATCTGCAATCACTGCAAGCACTA 342

Qy 1722 CCGAGCCCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1781
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 CAGCTGCAACTTACAACTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCT 402

Qy 1782 CTGCCCGCCAGCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1841
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 CAGCTTACCAACGACCAACAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCT 462

Qy 1842 AAATCCCAACTACTGCTACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 CAGCTGCAACTTACAACTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCT 522

Qy 1902 CCACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1924
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 CAGCTACCTCTTAAATTAATCA 545

RESULT 14
AU033659
LOCUS AU033659 603 bp mRNA EST 28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLB278, mRNA sequence.
ACCESSION AU033659
NID 93799083
KEYWORDS EST.
SOURCE Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
clone_lib:SL.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 603)
AUTHORS Urushihara,H.
TITLE Eukaryota; Dictyosteliida; Dictyostelium.
JOURNAL Developmental cDNA in Dictyostelium discoideum
REFERENCE 2 (bases 1 to 603)
AUTHORS Urushihara,H.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1998) to the DBJ/EMBL/GenBank databases. Hideko
Urushihara, University of Tsukuba, Institute of Biological
Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
(E-mail:gxurushi@bank.dna.afrc.go.jp, Tel:+81-0298-53-4664,
Fax:+81-0298-53-6614)
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
source
1..603
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SL"
/dev_stage="slug"
BASE COUNT 184 a 192 c 53 g 174 t
ORIGIN

Query Match          3.2%; Score 89.4; DB 31; Length 603;
Best Local Similarity 50.1%; Pred. No. 2.1e-08;
Matches 222; Conservative 0; Mismatches 221; Indels 0; Gaps 0;

Qy 1482 CATCGACTGCTGCCCGGACAGCTGCTCCAAAGGGTGCCTCCCAACGGCTGCCCAACGGCTG 1541
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 103 CATCTTCAACATCATCATCTGCTGCAAGCTCTTCTCCATCTTCCATCTGCTGCAAGCTCTT 162

Qy 1542 CCCCAACTGCCGACCTTCACCGTGTCCACCTGCAACGGCCAGCCGCAACTGCGAGCCC 1601
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 163 CTCATCTTTCATCTGCTGCAAGCTCTTCCCATCTTCCCATCTTCCATCTGCAAGCTCTTCCC 162

Qy 1602 CAGTCCCACTTACCAACCGCAATCTCTACCGGCGCCACCGCTGCCCGCCAGCGCTC 1661
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 223 CATCTTCACTGCAATCAAGCTCTGCCCATCTTCCATCTGCAATCAAGCTCTGCTC 282

Qy 1662 CTACTACTGCTGCCCGCTGAATCCCAACCACTGTGCAGTACCACTACTGCTGCCCA 1721
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 283 CATCTTCACTGCAATCACTTTCATCTGCAATCACTTTCATCTGCAATCACTGCAAGCACTA 342

Qy 1722 CCGAGCCCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1781
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 343 CAGCTGCAACTTACAACTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCT 402

Qy 1782 CTGCCCGCCAGCGTGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1841
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 403 CAGCTTACCAACGACCAACAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCT 462

Qy 1842 AAATCCCAACTACTGCTACATCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1901
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 463 CAGCTGCAACTTACAACTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCTGCAAGCT 522

Qy 1902 CCACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1924
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 523 CAGCTACCTCTTAAATTAATCA 545

RESULT 15
AU033708
LOCUS AU033708 586 bp mRNA EST 28-OCT-1998
DEFINITION Dictyostelium discoideum slug cDNA, clone SLB343, mRNA sequence.
ACCESSION AU033708
NID 93799132
KEYWORDS EST.
SOURCE Dictyostelium discoideum (strain:AX4) slug cDNA to mRNA,
clone_lib:SL.
ORGANISM Dictyostelium discoideum
REFERENCE 1 (bases 1 to 586)
AUTHORS Urushihara,H.
TITLE Eukaryota; Dictyosteliida; Dictyostelium.
JOURNAL Developmental cDNA in Dictyostelium discoideum
REFERENCE 2 (bases 1 to 586)
AUTHORS Urushihara,H.
TITLE Direct Submission
JOURNAL Submitted (22-SEP-1998) to the DBJ/EMBL/GenBank databases. Hideko
Urushihara, University of Tsukuba, Institute of Biological
Sciences; 1-1-1 Ten-noudai, Tsukuba-shi 305-8572, Japan
(E-mail:gxurushi@bank.dna.afrc.go.jp, Tel:+81-0298-53-4664,
Fax:+81-0298-53-6614)
PROJECT = 'Dictyostelium discoideum cDNA project in Japan'.
FEATURES
source
1..586
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone_lib="SL"
/dev_stage="slug"
BASE COUNT 182 a 186 c 50 g 168 t
ORIGIN

Query Match          3.2%; Score 89.4; DB 31; Length 586;
Best Local Similarity 50.1%; Pred. No. 2.1e-08;
Matches 222; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
```

```
Matches 222; Conservative 0; Mismatches 221; Indels 0; Gaps 0;
QY 1482 CATCGACTGCTGCCCGACAGCTGTCCAAAGGCTGCCCAAGGCTGCCCAAGGCTG 1541
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
26 CATCTTCACCATCATCATCTGTGTGCAAGCTCTTCCATCTTCATCTGTGTGCAAGCTCTT 85
QY 1542 CCGCAACTGCGGACCGCTCCAGGGTGTCCAGCTGCAAGCGCACCGCAACTGCAAGCCC 1601
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
86 CTCCATCTTTCATCTGTGTGCAAGCTCTTCCCATCTTTCATCTGTCATCATCAAGCTCTTCCC 145
QY 1602 CAGTCCACCTACAACCGCAATTCCTACTCGGCGCCCGCCCGCTGCCCGACGGAGCTC 1661
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
146 CATCTTCATCTGGATCATCAAGCTCTGCCCATCTTTCATCTGTCATCATCAAGCTCTGCTC 205
QY 1662 CTACTACTGCTGCCCTGAATCCCCAACCACTGTACAGTACCACTACTGTGTCTCCCA 1721
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
206 CATCTTCATCTGCATCATCTTTCATCTGCATCATCTTTCATCTGCATCATCTGCAGCAACTA 265
QY 1722 CCGCAGCCCTACTACTGTGTCCCTGAAATCCCAATCACTGTGCATCAGCGCTACCG 1781
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
266 CAGCTGCAACTACAATTGCAACTTACAGCTGCAACTACAAGTACAAGTACAAGCGCAACCA 325
QY 1782 CTGCCCCACCGCTGCCCGCACGGTGGCCCGCACCGCGCTACTGTGTCTCCAG 1841
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
326 CAGCTACAACCAAGCCACCAACCAAGCTACACCAAGCTGCAACCATAGCTACCAACCA 385
QY 1842 AAATCCCAACTACTGTACATCACCACCTACTGTGCCCGCCACTACCGCAGCACCTGCC 1901
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
386 CAGCTGCAACTACAAGTACAAGTACAAGTACAAGTACAAGTACAAGTACAAGTACAAGTACA 445
QY 1902 CCAACACCAAGTCACTGTACCA 1924
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
446 CAGCTACCTCTTAAATTTAATCA 468
```

Search completed: July 18, 1999, 11:36:45  
Job time: 1753 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 02:08:42 ; Search time 1514.49 Seconds  
(without alignments)  
4140.380 Million cell updates/sec

Title: US-09-294-663-2

Perfect score: 2821

Sequence: 1 GAAGAGATAACACGGAC.....TAAAAA.....2821

Scoring table: IDENTITY\_NUC

Searched: 3232953 seqs, 1111407957 residues

Database : Pending\_Patents\_NA.\*

1: /cgn2\_6/ptodata/1/pna/US0899A\_COMB.seq.\*  
2: /cgn2\_6/ptodata/1/pna/US07\_COMB.seq.\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq.\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq.\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq.\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq.\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq.\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq.\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq.\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq.\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq.\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq.\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq.\*  
14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq.\*  
15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq.\*  
16: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq.\*  
17: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq.\*  
18: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq.\*  
19: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq.\*  
20: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq.\*  
21: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq.\*  
22: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq.\*  
23: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq.\*  
24: /cgn2\_6/ptodata/1/pna/US088D\_COMB.seq.\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq.\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq.\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq.\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq.\*  
29: /cgn2\_6/ptodata/1/pna/US089E\_COMB.seq.\*  
30: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq.\*  
31: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq.\*  
32: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq.\*  
33: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq.\*  
34: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq.\*  
35: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq.\*  
36: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq.\*  
37: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq.\*  
38: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq.\*  
39: /cgn2\_6/ptodata/1/pna/US093\_COMB.seq.\*  
40: /cgn2\_6/ptodata/1/pna/US6000\_COMB.seq.\*  
41: /cgn2\_6/ptodata/1/pna/US6001\_COMB.seq.\*  
42: /cgn2\_6/ptodata/1/pna/US6002A\_COMB.seq.\*  
43: /cgn2\_6/ptodata/1/pna/US6002B\_COMB.seq.\*  
44: /cgn2\_6/ptodata/1/pna/US6003A\_COMB.seq.\*  
45: /cgn2\_6/ptodata/1/pna/US6003B\_COMB.seq.\*  
46: /cgn2\_6/ptodata/1/pna/US6004A\_COMB.seq.\*  
47: /cgn2\_6/ptodata/1/pna/US6004B\_COMB.seq.\*  
48: /cgn2\_6/ptodata/1/pna/US6005\_COMB.seq.\*  
49: /cgn2\_6/ptodata/1/pna/US6006\_COMB.seq.\*  
50: /cgn2\_6/ptodata/1/pna/US6007\_COMB.seq.\*  
51: /cgn2\_6/ptodata/1/pna/US6008A\_COMB.seq.\*  
52: /cgn2\_6/ptodata/1/pna/US6008B\_COMB.seq.\*  
53: /cgn2\_6/ptodata/1/pna/US6008C\_COMB.seq.\*

54: /cgn2\_6/ptodata/1/pna/US6009A\_COMB.seq.\*  
55: /cgn2\_6/ptodata/1/pna/US6009B\_COMB.seq.\*  
56: /cgn2\_6/ptodata/1/pna/US6010A\_COMB.seq.\*  
57: /cgn2\_6/ptodata/1/pna/US6010B\_COMB.seq.\*  
58: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq.\*  
59: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq.\*  
60: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq.\*  
61: /cgn2\_6/ptodata/1/pna/PCT\_NEW\_COMB.seq.\*  
62: /cgn2\_6/ptodata/1/pna/US07\_NEW\_COMB.seq.\*  
63: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq.\*  
64: /cgn2\_6/ptodata/1/pna/US09\_NEW\_COMB.seq.\*  
65: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2821	100.0	2821	33	US-09-103-429-2 Sequence 2, Appli
2	2821	100.0	2821	33	US-09-103-429A-2 Sequence 2, Appli
3	2821	100.0	2821	38	US-09-294-663-2 Sequence 2, Appli
4	2354.4	83.5	2455	33	US-09-103-429-1 Sequence 1, Appli
5	2354.4	83.5	2455	33	US-09-103-429A-1 Sequence 1, Appli
6	2354.4	83.5	2455	38	US-09-294-663-1 Sequence 1, Appli
7	158.8	5.6	8705	1	PCT-US99-06742-14 Sequence 14, Appli
8	158.8	5.6	10596	13	US-08-407-128-15 Sequence 15, Appli
9	158.8	5.6	10596	13	US-08-409-770-15 Sequence 15, Appli
10	158.8	5.6	10596	15	US-08-487-837-15 Sequence 15, Appli
11	158.8	5.6	10596	15	US-08-487-831-15 Sequence 15, Appli
12	158.8	5.6	803	18	US-08-733-369A-104 Sequence 104, App
13	158.8	5.6	799	18	US-08-733-369A-105 Sequence 105, App
14	158.8	5.6	798	18	US-08-733-369A-112 Sequence 112, App
15	158.8	5.6	9600	25	US-08-910-647-1 Sequence 1, Appli
16	158.8	5.6	798	28	US-08-970-900-4 Sequence 4, Appli
17	158.8	5.6	798	28	US-08-970-900-6 Sequence 6, Appli
18	158.8	5.6	803	28	US-08-970-900-7 Sequence 7, Appli
19	158.8	5.6	2580	31	US-09-050-863-2 Sequence 2, Appli
20	158.8	5.6	5452	33	US-09-130-114-1 Sequence 1, Appli
21	158.8	5.6	10380	33	US-09-130-115-1 Sequence 1, Appli
22	147	5.2	947	59	US-60-129-391-160 Sequence 160, App
23	137.8	4.9	5163	1	PCT-US97-14104-1 Sequence 1, Appli
24	137.8	4.9	5163	1	PCT-US97-14104-2 Sequence 1, Appli
25	137.8	4.9	5163	18	US-08-700-651A-1 Sequence 2, Appli
26	137.8	4.9	5163	18	US-08-700-651A-2 Sequence 2, Appli
27	137.8	4.9	5163	18	US-08-700-651-1 Sequence 1, Appli
28	137.8	4.9	5163	18	US-08-700-651-2 Sequence 2, Appli
29	137.8	4.9	5318	26	US-08-928-361-3 Sequence 3, Appli
30	137.8	4.9	5163	26	US-08-928-361-4 Sequence 4, Appli
31	137.8	4.9	5318	63	US-08-928-361B-3 Sequence 3, Appli
32	137.8	4.9	5163	63	US-08-928-361B-4 Sequence 4, Appli
33	135.2	4.8	7334	26	US-08-928-361-1 Sequence 1, Appli
34	135.2	4.8	5511	26	US-08-928-361-2 Sequence 2, Appli
35	135.2	4.8	7334	63	US-08-928-361B-1 Sequence 1, Appli
36	135.2	4.8	5511	63	US-08-928-361B-2 Sequence 2, Appli
37	120.6	4.3	565	60	US-60-138-103-14338 Sequence 14338, A
38	120.2	4.3	1010	60	US-60-138-103-13491 Sequence 13491, A
39	119.2	4.2	32207	1	PCT-US97-01442-20 Sequence 20, Appli
40	119.2	4.2	32207	1	PCT-US97-13346-20 Sequence 20, Appli
41	119.2	4.2	12754	16	US-08-592-963B-58 Sequence 58, Appli
42	119.2	4.2	12753	16	US-08-592-963B-59 Sequence 59, Appli
43	119.2	4.2	12752	16	US-08-592-963B-60 Sequence 60, Appli
44	119.2	4.2	12754	16	US-08-592-963B-61 Sequence 61, Appli
45	119.2	4.2	12753	16	US-08-592-963B-62 Sequence 62, Appli

ALIGNMENTS

RESULT 1

US-09-103-429-2  
; Sequence 2, Application US/09103429  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2821 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; DEVELOPMENTAL STAGE: larva  
; TISSUE TYPE: peritrophic membrane  
US-09-103-429-2

Query Match 100.0%; Score 2821; DB 33; Length 2821;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAAGAATAACC 13  
|||||  
Db 1 GAAAGAATAACC 13

RESULT 2  
US-09-103-429A-2  
; Sequence 2, Application US/09103429A  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850

; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2821 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; DEVELOPMENTAL STAGE: larva  
; TISSUE TYPE: peritrophic membrane  
US-09-103-429A-2

Query Match 100.0%; Score 2821; DB 33; Length 2821;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 13; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAAGAATAACC 13  
|||||  
Db 1 GAAAGAATAACC 13

RESULT 3  
US-09-294-663-2  
; Sequence 2, Application US/09294663  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioga Street  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/294,663  
; FILING DATE: 19-APR-1999  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/103,429  
; FILING DATE: 24-JUN-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390

REFERENCE/DOCKET NUMBER: BTL-39-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2821 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: not relevant  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
DEVELOPMENTAL STAGE: larva  
TISSUE TYPE: peritrophic membrane  
US-09-294-663-2

Query Match 100.0%; Score 2821; DB 38; Length 2821;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2821; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAAAAGATAACACGCGAACAAGTTATGATAAGACCCCTCTATTCTTGACGGCCCTCGG 60  
DB 1 GAAAAGATAACACGCGAACAAGTTATGATAAGACCCCTCTATTCTTGACGGCCCTCGG 60  
QY 61 GCTGCTGCGCGGCTCTGTAAGTCAGGACGCGAGAGAACCCGGCTCTCCAGAGCC 120  
DB 61 GCTGCTGCGCGGCTCTGTAAGTCAGGACGCGAGAGAACCCGGCTCTCCAGAGCC 120  
QY 121 GCACCCAGACTGCCCTCCGCTGAGCAGCAGTGGCTGCTGCTCAGTAAGACTGCAC 180  
DB 121 GCACCCAGACTGCCCTCCGCTGAGCAGCAGTGGCTGCTGCTCAGTAAGACTGCAC 180  
QY 181 CAAGTCTTACTACTGTGAATATGGTCTCAAGTTCATCGGACCGAGAGACTGTGCTCTGG 240  
DB 181 CAAGTCTTACTACTGTGAATATGGTCTCAAGTTCATCGGACCGAGAGACTGTGCTCTGG 240  
QY 241 TACCAGAAATCAAGTTCTCGCTCAGACTTGTGTTACGCGCGTTTGGCGGATGCACCT 300  
DB 241 TACCAGAAATCAAGTTCTCGCTCAGACTTGTGTTACGCGCGTTTGGCGGATGCACCT 300  
QY 301 GCCAGGACCTCCAGCTGAGACAAACCCAGGCCCCAGCAACAATCAGGCCCAACACAC 360  
DB 301 GCCAGGACCTCCAGCTGAGACAAACCCAGGCCCCAGCAACAATCAGGCCCAACACAC 360  
QY 361 CCAGGCGCCGCAACCACTACTAGGCGCCCTACTACACCCAGGCCCCAGGCCCCAACAC 420  
DB 361 CCAGGCGCCGCAACCACTACTAGGCGCCCTACTACACCCAGGCCCCAGGCCCCAACAC 420  
QY 421 CACCCAGGCCCCAACCAACCCAGGCCCCAACCAACCCAGGCCCCAACCACTACACCTCA 480  
DB 421 CACCCAGGCCCCAACCAACCCAGGCCCCAACCAACCCAGGCCCCAACCACTACACCTCA 480  
QY 481 GGCCCTTACTACTAGCACTCAGGCCCCAACCAACCACTCAGGCCCCCTTACCAACACAC 540  
DB 481 GGCCCTTACTACTAGCACTCAGGCCCCAACCAACCACTCAGGCCCCCTTACCAACACAC 540  
QY 541 CCAGGCGCCGCAACCAACCCAGGCCCCAACCACTACCAACCCAGGCCCCAACCACTCAGG 600  
DB 541 CCAGGCGCCGCAACCAACCCAGGCCCCAACCACTACCAACCCAGGCCCCAACCACTCAGG 600  
QY 601 CCCAACTACAATCAGGCTGCACTACCGGCGGCACTACCGGCGGCACTACCGGCGGCACTAC 660  
DB 601 CCCAACTACAATCAGGCTGCACTACCGGCGGCACTACCGGCGGCACTACCGGCGGCACTAC 660  
QY 661 CCGGCGGCACTACCGCTGCGGCGGCACTACCGGCGGCACTACCGGCGGCTGCACTACCGG 720  
DB 661 CCGGCGGCACTACCGCTGCGGCGGCACTACCGGCGGCACTACCGGCGGCTGCACTACCGG 720

QY 721 ACCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACCTGTGGCCAAATGGTTGCCCGAC 780  
DB 721 ACCACTTCAGCCCCAGTCTGGCCCCCGATCTGTGAACCTGTGGCCAAATGGTTGCCCGAC 780  
QY 781 TGACTTCGACATCCACTTGTGATTTCCCGCAGACAAAGTACTGCAACCTCTTCTACCACTG 840  
DB 781 TGACTTCGACATCCACTTGTGATTTCCCGCAGACAAAGTACTGCAACCTCTTCTACCACTG 840  
QY 841 CTCCAACGGTTACACCTTCGAAACAGAGGTGCCCTGAGGAGCTCTACTTCAACCCCTACGT 900  
DB 841 CTCCAACGGTTACACCTTCGAAACAGAGGTGCCCTGAGGAGCTCTACTTCAACCCCTACGT 900  
QY 901 CCAGGCTGCGACTCTCTCTGCTAAGCTTGAATGACGCGGCGGAAATCAGCCCGGACCC 960  
DB 901 CCAGGCTGCGACTCTCTCTGCTAAGCTTGAATGACGCGGCGGAAATCAGCCCGGACCC 960  
QY 961 AGTCACAGAAGGCAACGAGAGACATTGACATCGGAGACCTCTCGACAATGGATG 1020  
DB 961 AGTCACAGAAGGCAACGAGAGACATTGACATCGGAGACCTCTCGACAATGGATG 1020  
QY 1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTCCCCACCGAAACCGTTGCGACAATGATTA 1080  
DB 1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTCCCCACCGAAACCGTTGCGACAATGATTA 1080  
QY 1081 CCAGTGGCTCCACGGTAACTTGGTAGAGAGCGTGTGGAGCGGCGCACCTTCAGTTT 1140  
DB 1081 CCAGTGGCTCCACGGTAACTTGGTAGAGAGCGTGTGGAGCGGCGCACCTTCAGTTT 1140  
QY 1141 TGAATTCAGCAATGTGACACATCGAGCTGTGGCTGACCTTCCCGGCGGCGAGAG 1200  
DB 1141 TGAATTCAGCAATGTGACACATCGAGCTGTGGCTGACCTTCCCGGCGGCGAGAG 1200  
QY 1201 CGAAGAAGTTGACGTCGACGAGGATGCTTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260  
DB 1201 CGAAGAAGTTGACGTCGACGAGGATGCTTGCACCGGCTGGTACTGCCCCACGGAACCCAT 1260  
QY 1261 TGAATGGAGCGCTCCCGAACGCGTGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1320  
DB 1261 TGAATGGAGCGCTCCCGAACGCGTGGCTGCGGCTGCGGCTGCGGCTGCGGCTGCGG 1320  
QY 1321 CCACGAGAGGACTGCGGCGGCTGCTACAGTGTGCTGCTGACAGACTATCGCAAGACC 1380  
DB 1321 CCACGAGAGGACTGCGGCGGCTGCTACAGTGTGCTGCTGACAGACTATCGCAAGACC 1380  
QY 1381 TTGCCCTGAAAACCTTCAGTCTTCCGACACAGTCTGCTGAGTCTCTGTGACCGC 1440  
DB 1381 TTGCCCTGAAAACCTTCAGTCTTCCGACACAGTCTGCTGAGTCTCTGTGACCGC 1440  
QY 1441 TGGTTGCCAAGTTTTCGAGTGGGATTTGACAACAGTGCACATCGACTGTGCCCCGAC 1500  
DB 1441 TGGTTGCCAAGTTTTCGAGTGGGATTTGACAACAGTGCACATCGACTGTGCCCCGAC 1500  
QY 1501 AGCTGCTCAACGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGGCTGCC 1560  
DB 1501 AGCTGCTCAACGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGGCTGCC 1560  
QY 1561 CACCGTGTGCCACCTTCGACGCGCCACCGCAACTGCAGCCCCAGTCCCACTTACCAACCG 1620  
DB 1561 CACCGTGTGCCACCTTCGACGCGCCACCGCAACTGCAGCCCCAGTCCCACTTACCAACCG 1620  
QY 1621 AATTCCTACTCGGCCCCACGGTGGCCCCACCGAGCTCTTACTGTGCCCCCTGA 1680  
DB 1621 AATTCCTACTCGGCCCCACGGTGGCCCCACCGAGCTCTTACTGTGCCCCCTGA 1680  
QY 1681 ATCCCAACCACTGTCAGTACGACCTTACTGTCTCCGACCGGAGCCCTTACTACTGC 1740  
DB 1681 ATCCCAACCACTGTCAGTACGACCTTACTGTCTCCGACCGGAGCCCTTACTACTGC 1740  
QY 1741 TGTCCCTGAAATCCCAATCACTGTACATGAGGCTTACCGCTGCCCCACCGCTGCC 1800  
DB 1741 TGTCCCTGAAATCCCAATCACTGTACATGAGGCTTACCGCTGCCCCACCGCTGCC 1800  
QY 1801 CACCGCTGCCCGCCACCGGAGGCCCTTACTGTCTGCCAGAAATCCCACTTACTGT 1860



Db 1801 CACCGTGTCCCGCCAGCGAGCCCTACTACTGTGTGCCAGAAATCCCACTACTGTCTAC 1860  
QY 1861 ATCACCACCTACTGTCTGCCCGCCACTACCGCAGCAGCTGTGCCCGCCAAACACACAGTCTACTGT 1920  
Db 1861 ATCACCACCTACTGTCTGCCCGCCACTACCGCAGCAGCTGTGCCCGCCAAACACACAGTCTACTGT 1920  
QY 1921 ACCACCCACTGTCTGCCCGCCACTACTACCGCAGCAGCTGTGCCCGCCAAACACACAGTCTACTGT 1980  
Db 1921 ACCACCCACTGTCTGCCCGCCACTACTACCGCAGCAGCTGTGCCCGCCAAACACACAGTCTACTGT 1980  
QY 1981 ACCACCGCAGCGCCCTACTACTACCGCAGCAGCTGTGCCCGCCAAACACACAGTCTACTGTAC 2040  
Db 1981 ACCACCGCAGCGCCCTACTACTACCGCAGCAGCTGTGCCCGCCAAACACACAGTCTACTGTAC 2040  
QY 2041 CACTGTGTGCCCGCCACTGTACCTGTCCCTGTCCGCTGCAGCTGTCCCGCCAAACACACAGTCT 2100  
Db 2041 CACTGTGTGCCCGCCACTGTACCTGTCCCTGTCCGCTGCAGCTGTCCCGCCAAACACACAGTCT 2100  
QY 2101 CCGGTAACCTACACACGCGCAGCAGTACACACCTGTGAGATGTGACATCGACGCCGCC 2160  
Db 2101 CCGGTAACCTACACACGCGCAGCAGTACACACCTGTGAGATGTGACATCGACGCCGCC 2160  
QY 2161 TCTCCCAAGCAGCCCATCAACCCCTTGGCTTGAAGATGCAACGTTTGGCCCTGGGCTCA 2220  
Db 2161 TCTCCCAAGCAGCCCATCAACCCCTTGGCTTGAAGATGCAACGTTTGGCCCTGGGCTCA 2220  
QY 2221 CGTGTACTGGGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCTGTGTGTGTGTGTC 2280  
Db 2221 CGTGTACTGGGACAAATACTGGGTCTGTGACGGCAACAACCAAGTCTGTGTGTGTGTGTC 2280  
QY 2281 TGAGGTCTCCAGTTCAACCCCTACTTAAGACCTGTGACTTGCCTTGAACGTCGGTTG 2340  
Db 2281 TGAGGTCTCCAGTTCAACCCCTACTTAAGACCTGTGACTTGCCTTGAACGTCGGTTG 2340  
QY 2341 CGTGAGGAGCAACATTCAGATGTCTGAAGCTAGGAGGAGTCCAGTCTTCACTCCCATG 2400  
Db 2341 CGTGAGGAGCAACATTCAGATGTCTGAAGCTAGGAGGAGTCCAGTCTTCACTCCCATG 2400  
QY 2401 GAACAACTAGATGAAGACATCAGACGCGCTGAACCTTGTGAGTGTGAACCTACTTAA 2460  
Db 2401 GAACAACTAGATGAAGACATCAGACGCGCTGAACCTTGTGAGTGTGAACCTACTTAA 2460  
QY 2461 TTAATGAAGTTTGTGTTTATTTTTCAGTGTATTTTCAATGCGGGGAAAGTCCGCCAT 2520  
Db 2461 TTAATGAAGTTTGTGTTTATTTTTCAGTGTATTTTCAATGCGGGGAAAGTCCGCCAT 2520  
QY 2521 TATGGGTCTTGCAGTTTGGAGAACCTTTTTTTTACTACCAACATCTTGTGAACC 2580  
Db 2521 TATGGGTCTTGCAGTTTGGAGAACCTTTTTTTTACTACCAACATCTTGTGAACC 2580  
QY 2581 CATATTTATTTACGTATTAACATCGTATTTGAAACAGTTTACATGATTTTTTTCATTAAT 2640  
Db 2581 CATATTTATTTACGTATTAACATCGTATTTGAAACAGTTTACATGATTTTTTTCATTAAT 2640  
QY 2641 TTGAACAAATTTATGTTGTTTGTCTCATTAATATCAATATATTTTTCGAAACTGG 2700  
Db 2641 TTGAACAAATTTATGTTGTTTGTCTCATTAATATCAATATATTTTTCGAAACTGG 2700  
QY 2701 CAATTTGGATTTGAATATCAACAAATGTTTGAAGAAAAACGATTTCTTAAAAATGT 2760  
Db 2701 CAATTTGGATTTGAATATCAACAAATGTTTGAAGAAAAACGATTTCTTAAAAATGT 2760  
QY 2761 ATTTATTAATAATGTTAAATATAAATATCAAAATAGCATTTAAAAAATAAATAAATAA 2820  
Db 2761 ATTTATTAATAATGTTAAATATAAATATCAAAATAGCATTTAAAAAATAAATAAATAA 2820  
QY 2821 A 2821  
Db 2821 A 2821

RESULT 4

US-09-103-429-1  
; Sequence 1, Application US/09103429  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: cDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2455 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: Peritrophic Membrane  
; IMMEDIATE SOURCE:  
; CLONE: IIM14  
; US-09-103-429-1

Query Match 83.5%; Score 2354.4; DB 33; Length 2455;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;  
QY 1 GAAAGAGATAACACCGAACAAGTTATGATAAGACCTCTCTATTCCTGACGGCCCTCGG 60  
Db 13 GAAAGAGATAACACCGAACAAGTTATGATAAGACCTCTCTATTCCTGACGGCCCTCGG 72  
QY 61 GCTGTGCGCGCGCTCTGAGTTCAGCGACGCGGAGAGAACCCGCTCTCCACGAGCC 120  
Db 73 GCTGTGCGCGCGCTCTGAGTTCAGCGACGCGGAGAGAACCCGCTCTCCACGAGCC 132  
QY 121 GCACCCAGACTGCCCTCCGCTGAGCAGCAGTGGCTCTCGCTCAGGAATAGCAGTGCAC 180  
Db 133 GCACCCAGACTGCCCTCCGCTGAGCAGCAGTGGCTCTCGCTCAGGAATAGCAGTGCAC 192  
QY 181 CAAGTCTCTACTGTGAATATGGTCTCAAGTTCATCGCACCAGAGAGTGTCTCTCGG 240  
Db 193 CAAGTCTCTACTGTGAATATGGTCTCAAGTTCATCGCACCAGAGAGTGTCTCTCGG 252  
QY 241 TACCGAATTCAGTTCTCGCTCAGACTTGTGTTCACGCGGCTTTAGCCGGATGCACCT 300  
Db 253 TACCGAATTCAGTTCTCGCTCAGACTTGTGTTCACGCGGCTTTAGCCGGATGCACCT 312

QY 301 GCAGGACCTCAGCTGAGCAACCCAGGCCCGCCAGCAAACTCAGGCCCGCCAAACACAC 360  
 Db 313 GCCAGGACCTCAGCTGAGCAACCCAGGCCCGCCAGCAAACTCAGGCCCGCCAAACACAC 372  
 QY 361 CCAGGCCCGCCAAACCACTACTCAGGCCCGCTACTACAAACCCAGGCCCGCCAAACCAAC 420  
 Db 373 CCAGGCCCGCCAAACCACTACTCAGGCCCGCTACTACAAACCCAGGCCCGCCAAACCAAC 432  
 QY 421 CACCCAGGCCCGCCAAACCACTACTCAGGCCCGCTACTACAAACCCAGGCCCGCCAAACCAAC 480  
 Db 433 CACCCAGGCCCGCCAAACCACTACTCAGGCCCGCTACTACAAACCCAGGCCCGCCAAACCAAC 492  
 QY 481 GGCCTCTACTACTACCACTCAGGCCCGCCAGCAACCACTCAGGCCCGCTACTACCAACCAAC 540  
 Db 493 GGCCTCTACTACTACCACTCAGGCCCGCCAGCAACCACTCAGGCCCGCTACTACCAACCAAC 552  
 QY 541 CCAGGCCCGCCAAACCACTCAGGCCCGCCAGCAACCACTCAGGCCCGCTACTACCAACCAAC 600  
 Db 553 CCAGGCCCGCCAAACCACTCAGGCCCGCCAGCAACCACTCAGGCCCGCTACTACCAACCAAC 612  
 QY 601 CCCAACTACATCAGGCCCGCCAGCAACCACTCAGGCCCGCTACTACCAACCAACCAACCAAC 660  
 Db 613 CCCAACTACATCAGGCCCGCCAGCAACCACTCAGGCCCGCTACTACCAACCAACCAACCAAC 672  
 QY 661 CCGGCCCGCAACTACCCCTGCGGAGCAACCCCGCTCAGCACTACCCAGGCTGTTCTGTC 720  
 Db 673 CCGGCCCGCAACTACCCCTGCGGAGCAACCCCGCTCAGCACTACCCAGGCTGTTCTGTC 732  
 QY 721 ACCCACTCAGGCCCGCCAGCTGTCGCCCGCCAGCTGTCGAACCTGTTGCCCAATGTTGCCAGC 780  
 Db 733 ACCCACTCAGGCCCGCCAGCTGTCGCCCGCCAGCTGTCGAACCTGTTGCCCAATGTTGCCAGC 792  
 QY 781 TGACTTCGACATTCACCTTGTGATTCCCGGAGCAAGTACTGCAACCTCTTCTACCACTG 840  
 Db 793 TGACTTCGACATTCACCTTGTGATTCCCGGAGCAAGTACTGCAACCTCTTCTACCACTG 852  
 QY 841 CTCCAACGGTTACACCTTCGACAGAGTGCCTCAGGGAGCTCTACTTCAACCCCTAGCT 900  
 Db 853 CTCCAACGGTTACACCTTCGACAGAGTGCCTCAGGGAGCTCTACTTCAACCCCTAGCT 912  
 QY 901 CCAGGCTGCGACTCTCCTGCTAAGCTTGAATGCGAGCGGGAATCAGCCCGCCAGCCCGCC 960  
 Db 913 CCAGGCTGCGACTCTCCTGCTAAGCTTGAATGCGAGCGGGAATCAGCCCGCCAGCCCGCC 972  
 QY 961 AGTCAGAGGCAACGAGGAGCAAGACATTGACATCGGAGACCTCTCTGCAATGGATG 1020  
 Db 973 AGTCAGAGGCAACGAGGAGCAAGACATTGACATCGGAGACCTCTCTGCAATGGATG 1032  
 QY 1021 CCAGCTAAGCTTCGAAATCGACTGCTCTTGCCCGCCAGCAACCGTTGCGGACAGTATTA 1080  
 Db 1033 CCAGCTAAGCTTCGAAATCGACTGCTCTTGCCCGCCAGCAACCGTTGCGGACAGTATTA 1092  
 QY 1081 CCAGTGCCTCCAGGTAAGTGGTAGAGGGGTTGTTGGAGCGGCGACCTTCACTTCACTTT 1140  
 Db 1093 CCAGTGCCTCCAGGTAAGTGGTAGAGGGGTTGTTGGAGCGGCGACCTTCACTTCACTTT 1152  
 QY 1141 TGAATTCAGCAATGTACCACTGACATCGAGCTCGTTGGCTGCACTCCCGCGGCGAGAG 1200  
 Db 1153 TGAATTCAGCAATGTACCACTGACATCGAGCTCGTTGGCTGCACTCCCGCGGCGAGAG 1212  
 QY 1201 CGAAGAGTTGACCTGCGAGGAGGATGCTTGCACCGGCTGGTACTGCGCCCGCGGAGCCCAT 1260  
 Db 1213 CGAAGAGTTGACCTGCGAGGAGGATGCTTGCACCGGCTGGTACTGCGCCCGCGGAGCCCAT 1272  
 QY 1261 TGAATGGAGCGCCCTCCCGCAACGGCTGCGCTGCGGCTTACGATTCAGCATCGACCTCTCC 1320  
 Db 1273 TGAATGGAGCGCCCTCCCGCAACGGCTGCGCTGCGGCTTACGATTCAGCATCGACCTCTCC 1332  
 QY 1321 CCACGAGGCGACTCGCGCCAGTATCTACAGTGTCTCATGGACAGACTATCGCAAGACC 1380  
 Db 1333 CCACGAGGCGACTCGCGCCAGTATCTACAGTGTCTCATGGACAGACTATCGCAAGACC 1392  
 QY 1381 TTGCCCTGGAAACCTGCACTTACGTCCTGCCACACAGTCTCTGTGAGTCTCTGTGACCGC 1440

Db 1393 TTGCCCTGGAAACCTGCACTTCACTTCTGTCACAGTCTCTGTGAGTCTCTGTGACCGC 1452  
 QY 1441 TGGTTGCCAAGTTTTTCAGTGCAGTTTGTGACAAACAGTGCACATCGACTGCTGCCCCGAC 1500  
 Db 1453 TGGTTGCCAAGTTTTTCAGTGCAGTTTGTGACAAACAGTGCACATCGACTGCTGCCCCGAC 1512  
 QY 1501 AGTGTCTCCAAAGGCTGCCCCAAACGGTGCCTCCAAAGGCTGCCCAACTGCGCGACGCTC 1560  
 Db 1513 AGTGTCTCCAAAGGCTGCCCCAAACGGTGCCTCCAAAGGCTGCCCAACTGCGCGACGCTC 1572  
 QY 1561 CACCGTGTCCCACTGCAAGCGCCACCGCAACTGCAAGCGCCAGTGCAGCGCCAGTGCACCAACCGC 1620  
 Db 1573 CACCGTGTCCCACTGCAAGCGCCACCGCAACTGCAAGCGCCAGTGCAGCGCCAGTGCACCAACCGC 1632  
 QY 1621 AATTCTTACTTCCGCGCCCGCCAGCGTGCCTCCCGCCAGCGCTTCTACTTCTGCTGCCCCCTGA 1680  
 Db 1633 AATTCTTACTTCCGCGCCCGCCAGCGTGCCTCCCGCCAGCGCTTCTACTTCTGCTGCCCCCTGA 1692  
 QY 1681 ATCCCCAAACCACTGTCAGTACCACTTACTGCTGCTCCCGCCAGCGAGCCCTACTACTGTC 1740  
 Db 1693 ATCCCCAAACCACTGTCAGTACCACTTACTGCTGCTCCCGCCAGCGAGCCCTACTACTGTC 1752  
 QY 1741 TGTCCCTGAATCCCAATCACTGTCACATCAGCGCTACCGCTGCCCGCCAGCGCTGCCCC 1800  
 Db 1753 TGTCCCTGAATCCCAATCACTGTCACATCAGCGCTACCGCTGCCCGCCAGCGCTGCCCC 1812  
 QY 1801 CACCGCTGCCCGCCCGCCAGCGCTTACTTCTGCTCCCGCAAAATCCCAACTTACTGTCAC 1860  
 Db 1813 CACCGCTGCCCGCCCGCCAGCGCTTACTTCTGCTCCCGCAAAATCCCAACTTACTGTCAC 1872  
 QY 1861 ATCACCACCTTACTGCTGCCCGCCAGCGAGCAGCTGCCCCCAACACACACAGTCACTGT 1920  
 Db 1873 ATCACCACCTTACTGCTGCCCGCCAGCTGCGGAGCAGCTGCCCCCAACACACAGTCACTGT 1932  
 QY 1921 ACCACCCACTGCTGCCCGCTTACTACCGGAGCAGCTGCTCCCGCAACACACAGTCACTGAC 1980  
 Db 1933 ACCACCCACTGCT----- 1946  
 QY 1981 ACCCACCGCAGCCCTTACTACCGCAGCAGCTGCCCGCAACACACAGTCACTGTACCAC 2040  
 Db 1946 -----GCCCTTACTACCGCAGCAGCTGCCCGCAACACACAGTCACTGTACCAC 1995  
 QY 2041 CACTGCTGCCCGCCACTGAGCTCCCGCTACCGTGCACCTGCGACCTCCACACACACAGTGC 2100  
 Db 1996 CACTGCTGCCCGCCACTGAGCTCCCGCTACCGTGCACATGCAACCAACACACAGTGC 2055  
 QY 2101 CCGGTTAACTACACACGCGCAGCTACCACTGCAACCTGGAAGATGATGACATCGACCGCC 2160  
 Db 2056 CCGGTTAACTACACACGCGCAGCTACCACTGCAACCTGGAAGATGATGACATCGACCGCC 2115  
 QY 2161 TCTCCCAACGACCCCTCAACCTTGGCTTGAAGATGCAACCTTTTCCCTGGGCTCA 2220  
 Db 2116 TCTCCCAACGACCCCTCAACCTTGGCTTGAAGATGCAACCTTTTCCCTGGGCTCA 2175  
 QY 2221 CGCTGACTGCGCAAAATACTGGGCTGTGAGGCAACCAACCAAGTCTCTGGTGGTTC 2280  
 Db 2176 CGCTGACTGCGCAAAATACTGGGCTGTGAGGCAACCAACCAAGTCTCTGGTGGTTC 2235  
 QY 2281 TGAGGCTTCCAGTTCAACCCCACTACTAGACCTGTGACTTCGCTTGCACAGCTCGGTTG 2340  
 Db 2236 TGAGGCTTCCAGTTCAACCCCACTACTAGACCTGTGACTTCGCTTGCACAGCTCGGTTG 2295  
 QY 2341 CGTGAGGAGCAACATTCAGATGCTGAAAGCTACGAAGGAGTCCAGGCTTCTCATCCCATG 2400  
 Db 2296 CGTGAGGAGCAACATTCAGATGCTGAAAGCTACGAAGGAGTCCAGGCTTCTCATCCCATG 2355  
 QY 2401 GAACAACTAGATGAAGACATCAGACGCGCTGAACCTTTGAGTTGTAACCTACTTAAA 2460  
 Db 2356 GAACAACTAGATGAAGACATCAGACGCGCTGAACCTTTGAGTTGTAACCTACTTAAA 2415  
 QY 2461 TTAATGAAGGTTTGTGTTT 2480  
 Db 2415 TTAATGAAGGTTTGTGTTT

Db 2416 TTAATGAAGGTTTGTGTTTA 2435

## RESULT 5

US-09-103-429A-1  
; Sequence 1, Application US/09103429A  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: cDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaelis, P.C.  
; STREET: 118 North Tioga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09103,429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2455 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: Peritrophic Membrane  
; IMMEDIATE SOURCE:  
; CLONE: IIM14  
US-09-103-429A-1

Query Match 83.5%; Score 2354.4; DB 33; Length 2455;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;

QY 1 GAAAGAAATACCAGGGAACAAGTTATGATAAGACCCCTCTATTCTGACGGCCCTCGG 60  
Db 13 GAAAGAAATACCAGGGAACAAGTTATGATAAGACCCCTCTATTCTGACGGCCCTCGG 72  
QY 61 GCTGTCGCGCGGCGTCTGAAAGTACGACGCGGAGAGAACCCCGCTCTCCACGAGCC 120  
Db 73 GCTGTCGCGCGGCGTCTGAAAGTACGACGCGGAGAGAACCCCGCTCTCCACGAGCC 132  
QY 121 GCACCCAGACTGCCCTCCGCTGAGCAGCAGCTGGCTCTCCCTCAGCAATACGACTGCAC 180  
Db 133 GCACCCAGACTGCCCTCCGCTGAGCAGCAGCTGGCTCTCCCTCAGCAATACGACTGCAC 192  
QY 181 CAAGTCTTACTACTGTGAATATGGTCTCAAGTTCATCGCACGAGAGACTGTGCTCCTGG 240  
Db 193 CAAGTCTTACTACTGTGAATATGGTCTCAAGTTCATCGCACGAGAGACTGTGCTCCTGG 252

QY 241 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTACCGCCGCTTTAGCGGGATGACCCCT 300  
Db 253 TACCGAATTCAAGTTCTCCGCTCAGACTTGTGTTACCGCCGCTTTAGCGGGATGACCCCT 312  
QY 301 GCCAGGACCTCCAGCTGAGACAAACCCAGGCCCCAGCACAACACTCAGGCCGCCAACACAC 360  
Db 313 GCCAGGACCTCCAGCTGAGACAAACCCAGGCCCCAGCACAACACTCAGGCCGCCAACACAC 372  
QY 361 CCAGGCCCCAAACCAACTACTCAGGCCCTTACTACAAACCCAGGCCGCCAACACACAC 420  
Db 373 CCAGGCCCCAAACCAACTACTCAGGCCCTTACTACAAACCCAGGCCGCCAACACACAC 432  
QY 421 CACCCAGGCCCCAAACCAACTACTCAGGCCCTTACTACAAACCCAGGCCGCCAACACAC 480  
Db 433 CACCCAGGCCCCAAACCAACTACTCAGGCCCTTACTACAAACCCAGGCCGCCAACACAC 492  
QY 481 GGCCTTACTACTACCACTCAGGCCCTTACTACAAACCCAGGCCCTTACTACCAACAC 540  
Db 493 GGCCTTACTACTACCACTCAGGCCCTTACTACAAACCCAGGCCCTTACTACCAACAC 552  
QY 541 CCAGGCCCCAAACCAACTACTCAGGCCCTTACTACAAACCCAGGCCCTTACTACCAACAC 600  
Db 553 CCAGGCCCCAAACCAACTACTCAGGCCCTTACTACAAACCCAGGCCCTTACTACCAACAC 612  
QY 601 CCCAACTACAATCACCAGGCTGCAACTACCCGCGCCGCAACTACCCGCGCCGCAACTAC 660  
Db 613 CCCAACTACAATCACCAGGCTGCAACTACCCGCGCCGCAACTACCCGCGCCGCAACTAC 672  
QY 661 CCCGCGCCGCAACTACCCCTGCGCGGCAACCCCGCTGCAACTACCCGCGCTTCTTCGC 720  
Db 673 CCCGCGCCGCAACTACCCCTGCGCGGCAACCCCGCTGCAACTACCCGCGCTTCTTCGC 732  
QY 721 ACCCACTTCAGCCCTGCTGCGCCCGCATCTGTGAACCTGTGCCCAATGGTGGCCACG 780  
Db 733 ACCCACTTCAGCCCTGCTGCGCCCGCATCTGTGAACCTGTGCCCAATGGTGGCCACG 792  
QY 781 TGACTTCGACATCCACTTGTGTTGATTCGCCAGACAGACTGCAACCTCTTACCAGTG 840  
Db 793 TGACTTCGACATCCACTTGTGTTGATTCGCCAGACAGACTGCAACCTCTTACCAGTG 852  
QY 841 CTCACACGTTTACACCTTCGACAGAGGTGCGCTGAGGAGCTCTACTTCAACCCCTAGT 900  
Db 853 CTCACACGTTTACACCTTCGACAGAGGTGCGCTGAGGAGCTCTACTTCAACCCCTAGT 912  
QY 901 CCAGCGCTCGCACTCTCCTGCTTAAGTTGCAATGCGACGCGGCAAAATCAGCCCGCAC 960  
Db 913 CCAGCGCTCGCACTCTCCTGCTTAAGTTGCAATGCGACGCGGCAAAATCAGCCCGCAC 972  
QY 961 AGTCACAGAGGCAACGAGAGAGAGACATTGACATCGGAGACCTCTCGACAAATGGATG 1020  
Db 973 AGTCACAGAGGCAACGAGAGAGAGACATTGACATCGGAGACCTCTCGACAAATGGATG 1032  
QY 1021 CCCAGCTAACTTCGAAATCGACTGGCTCTTGGCCCAACGAAACCGTTGGCAAGATATTA 1080  
Db 1033 CCCAGCTAACTTCGAAATCGACTGGCTCTTGGCCCAACGAAACCGTTGGCAAGATATTA 1092  
QY 1081 CCAGTGCCTCCACGCTAACTTGGTAGAGAGCGTGTGGAGCGCGCACCTTCTCAGTTT 1140  
Db 1093 CCAGTGCCTCCACGCTAACTTGGTAGAGAGCGTGTGGAGCGCGCACCTTCTCAGTTT 1152  
QY 1141 TGAATTCAGCAATGTACCAATCAGCTGCTGGCTGACACCTCTCCCGGGCGGAGAG 1200  
Db 1153 TGAATTCAGCAATGTACCAATCAGCTGCTGGCTGACACCTCTCCCGGGCGGAGAG 1212  
QY 1201 CGAAGAACTTGCAGTCGAGGAGTGCCTGACCGGCTGGTACTGCCCGCACGGAACCCAT 1260  
Db 1213 CGAAGAACTTGCAGTCGAGGAGTGCCTGACCGGCTGGTACTGCCCGCACGGAACCCAT 1272  
QY 1261 TGAATGGGAGCCCTCCCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1320  
Db 1273 TGAATGGGAGCCCTCCCGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1332  
QY 1321 CCACGAGAGCGACTCGCGCCAGTATCTACAGTGTGTCCATGGACAGACTATCGCAAGACC 1380

Db 1333 CCACGAGAGGACTGGCGCAGTATCTACAGTGTCTCAATGACACACTATCCAAAGACC 1392  
QY 1381 TTGCCCTGGAACCTGACCTTACAGTCTGCCACACAGTCTCTGTAGTCTCTCTGTGACCGC 1440  
Db 1393 TTGCCCTGGAACCTGACCTTACAGTCTGCCACACAGTCTCTGTAGTCTCTCTGTGACCGC 1452  
QY 1441 TGGTTGCCAAGTTTTCAGATGCGAATCTGACAAACAGTGCATGCTGCTGCCCGAC 1500  
Db 1453 TGGTTGCCAAGTTTTCAGATGCGAATCTGACAAACAGTGCATGCTGCTGCCCGAC 1512  
QY 1501 AGTGTCTCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTC 1560  
Db 1513 AGTGTCTCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTGCCCAACGGCTC 1572  
QY 1561 CACCGTGGTCCCACTGCAACGGCCACCGCACTGACGAGCCCACTGACGAGCCCACTGACGAGCC 1620  
Db 1573 CACCGTGGTCCCACTGCAACGGCCACCGCACTGACGAGCCCACTGACGAGCCCACTGACGAGCC 1632  
QY 1621 AATTCCTACTCCGGCCCGCCACCGCTGCCCGCCACCGAGCTCTCTACTGCTGCTGCCCTGA 1680  
Db 1633 AATTCCTACTCCGGCCCGCCACCGCTGCCCGCCACCGAGCTCTCTACTGCTGCTGCCCTGA 1692  
QY 1681 ATCCCAACCACTGTACAGTACACCTACTGCTGCTCCACCGAGCCCGCTACTACTGC 1740  
Db 1693 ATCCCAACCACTGTACAGTACACCTACTGCTGCTCCACCGAGCCCGCTACTACTGC 1752  
QY 1741 TGTCCCTGAAATCCCAATCACTGTACATCAGCGCCTACCGTGCCTCCCGCCACCGCTGCCCG 1800  
Db 1753 TGTCCCTGAAATCCCAATCACTGTACATCAGCGCCTACCGTGCCTCCCGCCACCGCTGCCCG 1812  
QY 1801 CACCGTGCCTCCCGCCACCGAGCCCGCTACTGCTGCTCCAGAAATCCCACTACTGCTAC 1860  
Db 1813 CACCGTGCCTCCCGCCACCGAGCCCGCTACTGCTGCTCCAGAAATCCCACTACTGCTAC 1872  
QY 1861 ATCACCACCTACTGCTGCCCGCCACCGAGCAGCTGCTGCTCCAGAAATCCCACTACTGCT 1920  
Db 1873 ATCACCACCTACTGCTGCCCGCCACCGAGCAGCAGCTGCTGCTCCAGAAATCCCACTACTG 1932  
QY 1921 ACCACCACCTACTGCTGCCCGCCACCGAGCAGCAGCTGCTGCTCCAGAAATCCCACTACTG 1980  
Db 1933 ACCACCACCTACTGCTGCCCGCCACCGAGCAGCAGCTGCTGCTCCAGAAATCCCACTACTG 1946  
QY 1981 ACCACCAGGAGCCCGCTACTACCGAGCAGCAGCTGCTGCTCCAGAAATCCCACTACTGCT 2040  
Db 1946 -----GCCCTACTACCGAGCAGCAGCTGCTGCTCCAGAAATCCCACTACTGCTAC 1995  
QY 2041 CACTGCTGCCCGCCACCTGACGCTCCCGCTACCGTGCAGCTGCTGCTCCAGAAATCCCACTACTG 2100  
Db 1996 CACTGCTGCCCGCCACCTGACGCTCCCGCTACCGTGCAGCTGCTGCTCCAGAAATCCCACTACTG 2055  
QY 2101 CCGGTAAGTACAAACGAGCGCAGCTACCACTGAGATGATGATGATGATGATGATGATGATGATG 2160  
Db 2056 CCGGTAAGTACAAACGAGCGCAGCTACCACTGAGATGATGATGATGATGATGATGATGATGATG 2115  
QY 2161 TCTCCCAAGAGCCCGCATCAACCTTGGTGTGAAGATGAAGATGATGATGATGATGATGATGATG 2220  
Db 2116 TCTCCCAAGAGCCCGCATCAACCTTGGTGTGAAGATGAAGATGATGATGATGATGATGATGATG 2175  
QY 2221 CGCTGACTGGCAATATCTGGTGTCTGACGGCAGCAACCAAGTCTGCTGCTGCTGCTGCTGCTG 2280  
Db 2176 CGCTGACTGGCAATATCTGGTGTCTGACGGCAGCAACCAAGTCTGCTGCTGCTGCTGCTGCTG 2235  
QY 2281 TGAGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTGCCTTGCAACGCTGCTGCTGCTG 2340  
Db 2236 TGAGGTCTCCAGTTCAACCCCACTACTAAGACCTGTGACTTGCCTTGCAACGCTGCTGCTGCTG 2295  
QY 2341 CGTGAGAGCAACATTCAGATGCTGAAAGCTACGAAGAGTCCAGGTCTTCAATCCCATG 2400  
Db 2296 CGTGAGAGCAACATTCAGATGCTGAAAGCTACGAAGAGTCCAGGTCTTCAATCCCATG 2355  
QY 2401 GAACAACCTAGTAGACATCAGACGCGCTGAACTTTGAGTTGATGTTGATGTTGATGTTGATGTTG 2460

Db 2356 GAACAACCTAGATGAGACATCAGACAGCGCTGAACCTTTGAGTTGTTGTTGTTGTTGTTGTTG 2415  
QY 2461 TTAATGAAGTTTGTGTTT 2480  
Db 2416 TTAATGAAGTTTGTGTTT 2435  
RESULT 6  
US-09-294-663-1  
; Sequence 1, Application US/09294663  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tloga Street  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/294,663  
; FILING DATE: 19-APR-1999  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 09/103,429  
; FILING DATE: 24-JUN-1998  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39-CIP  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2455 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: Peritrophic Membrane  
; IMMEDIATE SOURCE:  
; CLONE: IIM14  
US-09-294-663-1  
Query Match 83.5%; Score 2354.4; DB 38; Length 2455;  
Best Local Similarity 97.7%; Pred. NO. 0;  
Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;  
QY 1 GAAAGAAATAAACCGGCAAGTATGATAAGACCCCTCTATTCCTGACGGCCCTCG 60  
Db 13 GAAAGAAATAAACCGGCAAGTATGATAAGACCCCTCTATTCCTGACGGCCCTCG 72  
QY 61 GCTCGTGGCGCGCTCTGAAGTACGAGCAGCGGAGAGAACCCCGCTCTCCACGAGCC 120  
Db 73 GCTCGTGGCGCGCTCTGAAGTACGAGCAGCGGAGAGAACCCCGCTCTCCACGAGCC 132  
QY 121 GCACCCAGACTGCCCTCCCGCTGAGCAGCACTGGCTCTGCTCAGTAAGTACTGCAC 180

Db 133 GCACCCAGACTGCCCTCCCGGTGAGCAGCACTGGCTCCTGCCTCACGAATACGACTGCAC 192  
QY 181 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTTCATCGCACCCAGAGAGACTGTGCTCCTGG 240  
Db 193 CAAGTTCTACTACTGTGAATATGGTCTCAAGTTTCATCGCACCCAGAGAGACTGTGCTCCTGG 252  
QY 241 TACCGAATTCAAGTTCTCCCGCTCAGACTTGTGTTCAGCCGCGCTTTAGCCGGATGCAACCT 300  
Db 253 TACCGAATTCAAGTTCTCCCGCTCAGACTTGTGTTCAGCCGCGCTTTAGCCGGATGCAACCT 312  
QY 301 GCCAGGACCTCCAGCTGAGACACCCAGGCCCGCAGCAACAACTCAGGCCCGCCCAACACCAC 360  
Db 313 GCCAGGACCTCCAGCTGAGACACCCAGGCCCGCAGCAACAACTCAGGCCCGCCCAACACCAC 372  
QY 361 CCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACAAACCAACCCAGCCCGCCCAACCAAC 420  
Db 373 CCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACAAACCAACCCAGCCCGCCCAACCAAC 432  
QY 421 CACCCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACAAACCAACCCAGCCCGCCCAACCAAC 480  
Db 433 CACCCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACAAACCAACCCAGCCCGCCCAACCAAC 492  
QY 481 GGCCCTACTACTACCACTCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACCAACCAACCC 540  
Db 493 GGCCCTACTACTACCACTCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACCAACCAACCC 552  
QY 541 CCAGGCCCGCCCAACCACTACTCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACCAACCAACCC 600  
Db 553 CCAGGCCCGCCCAACCACTACTCAGGCCCGCCCAACCACTACTCAGGCCCGCTACTACCAACCAACCC 612  
QY 601 CGCAACTACAACTACCCAGGCTGCACTACCCGGCGCGCAACTACCCGGCGCGCAACTAC 660  
Db 613 CCCAACTACAACTACCCAGGCTGCACTACCCGGCGCGCAACTACCCGGCGCGCAACTAC 672  
QY 661 CCGGCGCGCAACTACCCCTGCGCGGAGAACCCCGCTGCAACTACCCAGGCTGCTCTGCTGC 720  
Db 673 CCGGCGCGCAACTACCCCTGCGCGGAGAACCCCGCTGCAACTACCCAGGCTGCTCTGCTGC 732  
QY 721 ACCCACTACAGCCCGCACTGCGCCCGCGATCTGTGAACCTGTGCCCCAATGGTTGCCCGAGC 780  
Db 733 ACCCACTACAGCCCGCACTGCGCCCGCGATCTGTGAACCTGTGCCCCAATGGTTGCCCGAGC 792  
QY 781 TGACTTCGACATCCACTGTGTGATTCGCCAGCAAGTACTGCAACCTCTTTCACCACTG 840  
Db 793 TGACTTCGACATCCACTGTGTGATTCGCCAGCAAGTACTGCAACCTCTTTCACCACTG 852  
QY 841 CTCACCGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 900  
Db 853 CTCACCGGTTACACCTTCGAACAGAGGTGCCCTGAGGGACTCTACTTCAACCCCTACGT 912  
QY 901 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGCGCGAATACAGCCCGCGCACCCCC 960  
Db 913 CCAGCGCTGCGACTCTCCTGCTAACGTTGAATGCGCGCGGGAATACAGCCCGCGCACCCCC 972  
QY 961 AGTCACAGAAGGCAACGAAGCAAGACATGTGATCGGAGACCTCTCTCGACAATGGATG 1020  
Db 973 AGTCACAGAAGGCAACGAAGCAAGACATGTGATCGGAGACCTCTCTCGACAATGGATG 1032  
QY 1021 CCCAGCTTAACCTCGAATCCAGCTGGCTCTTGGCCCGGGAACCGTTGCGACAGTATTA 1080  
Db 1033 CCCAGCTTAACCTCGAATCCAGCTGGCTCTTGGCCCGGGAACCGTTGCGACAGTATTA 1092  
QY 1081 CCAGTGGCTCCAGGTAACCTTGGTAGAGAGCGGTTGTGAGCGCGGCAACCCACTTCAGTTT 1140  
Db 1093 CCAGTGGCTCCAGGTAACCTTGGTAGAGAGCGGTTGTGAGCGCGGCAACCCACTTCAGTTT 1152  
QY 1141 TGAACCTCAGCAATGTGACCACTCGAGCTCGTTGGCTGCAACCCCTCCCGCGCGCGAGAG 1200  
Db 1153 TGAACCTCAGCAATGTGACCACTCGAGCTCGTTGGCTGCAACCCCTCCCGCGCGCGAGAG 1212  
QY 1201 CGAAGAAGTTGACCTCCACGAGGATGCTGCACCGGCTGCTACTGCCCGCGAGTCC 1260

Db 1213 CGAAGAAGTTGAGCTCGACGAGGATGCTGCAACCGGCTGGTACTGCCCGCACGAACCCAT 1272  
QY 1261 TGAATGGGAGCCCTCCCAACAGGGTGCCTGCCGACTTCAGCATCGACCACTCCTCCTCC 1320  
Db 1273 TGAATGGGAGCCCTCCCAACAGGGTGCCTGCCGACTTCAGCATCGACCACTCCTCCTCC 1332  
QY 1321 CCAGGAGCGGACTGGGGCAGTATCTACAGTGTGTCCATGAGACACTATCCGACAGCC 1380  
Db 1333 CCAGGAGCGGACTGGGGCAGTATCTACAGTGTGTCCATGAGACACTATCCGACAGCC 1392  
QY 1381 TTGSCCTCGAAACCTCAGCTTCACTGCTGCAACAGTCTGCTGAGTCTCCTGTGACCGC 1440  
Db 1393 TTGSCCTCGAAACCTCAGCTTCACTGCTGCAACAGTCTGCTGAGTCTCCTGTGACCGC 1452  
QY 1441 TGTGTTCCCAAGTTTTCGAGTGGGATTTGACAAACAGTGCACATCGACTGCTGCCCGAC 1500  
Db 1453 TGTGTTCCCAAGTTTTCGAGTGGGATTTGACAAACAGTGCACATCGACTGCTGCCCGAC 1512  
QY 1501 AGTGTCTCCAAAGGGTGGCCCAACGGCTGCCCAACGGTGGCCCAACTGCGCGACCCCTC 1560  
Db 1513 AGTGTCTCCAAAGGGTGGCCCAACGGCTGCCCAACGGTGGCCCAACTGCGCGACCCCTC 1572  
QY 1561 CACCGTGTGCCCACTGCAAGCCACCCGCACTGCAAGCCCGCAGTCCCACTCAACACCGC 1620  
Db 1573 CACCGTGTGCCCACTGCAAGCCACCCGCACTGCAAGCCCGCAGTCCCACTCAACACCGC 1632  
QY 1621 AATTCCTACTCGGGCGCCCGCACCGCTGCCCGCACCGAGTCTCTACTACTGCTGCCCTGA 1680  
Db 1633 AATTCCTACTCGGGCGCCCGCACCGCTGCCCGCACCGAGTCTCTACTACTGCTGCCCTGA 1692  
QY 1681 ATCCCAACCACTGTCCAGTACCACCTACTGTGTCTCCACCGGAGCCCTACTACTGTC 1740  
Db 1693 ATCCCAACCACTGTCCAGTACCACCTACTGTGTCTCCACCGGAGCCCTACTACTGTC 1752  
QY 1741 TGTCCCTGAAATCCCAATCAGTGTACATCAGCGCTACCGTGCCTGCCCGCACCGCTGCC 1800  
Db 1753 TGTCCCTGAAATCCCAATCAGTGTACATCAGCGCTACCGTGCCTGCCCGCACCGCTGCC 1812  
QY 1801 CACCGTGGCCCGCACCGCGCCCTACTACTGTGTCTCCAGAAATCCCACTACTGTAC 1860  
Db 1813 CACCGTGGCCCGCACCGCGCCCTACTACTGTGTCTCCAGAAATCCCACTACTGTAC 1872  
QY 1861 ATCACCACTACTGTGTGCCCGCACTACCGGAGCACTGTGCCCGCAACACCACTACTGT 1920  
Db 1873 ATCACCACTACTGTGTGCCCGCACTACCGGAGCACTGTGCCCGCAACACCACTACTGT 1932  
QY 1921 ACCACCACTGTGTGCCCTACTTACCGCAGCACTGTCTCCCAACACCACTACTGTAC 1980  
Db 1933 ACCACCACTGTGT----- 1946  
QY 1981 ACCACCGGAGCCCTTACTACCGCAGCACTGTGCCCGCAACACCACTACTGTACCA 2040  
Db 1946 -----GCCCTTACTACCGCAGCACTGTGCCCGCAACACCACTACTGTACCA 1995  
QY 2041 CACTGTGCCCGCACTGTGAGTGTGCCCTTACCGTGGCACTGTGCCCGCAACACCACTACTGT 2100  
Db 1996 CACTGTGCCCGCACTGTGAGTGTGCCCTTACCGTGGCACTGTGCCCGCAACACCACTACTGT 2055  
QY 2101 CCGGTTAACTACAAACCAAGCGCACTTACCACTTACCACTTGAAGATGATGATGATGATGATG 2160  
Db 2056 CCGGTTAACTACAAACCAAGCGCACTTACCACTTACCACTTGAAGATGATGATGATGATGATG 2115  
QY 2161 TCTCCCAACGAGCCCATCAACCTTTCGCTTGAAGATGATGATGATGATGATGATGATGATG 2220  
Db 2116 TCTCCCAACGAGCCCATCAACCTTTCGCTTGAAGATGATGATGATGATGATGATGATGATG 2175  
QY 2221 CGCTGACTGGGACAAATACTGGGTCTGTGACGCAACCAACCACTTCTGGTGGTGGTGGTGGT 2280  
Db 2176 CGCTGACTGGGACAAATACTGGGTCTGTGACGCAACCAACCACTTCTGGTGGTGGTGGTGGT 2235  
QY 2281 TGAGGTCTCCAGTTCAACCCCACTACTTAAGACCTGTGACTTTCGTTGCAACCTGCGTTG 2340  
Db 2236 TGAGGTCTCCAGTTCAACCCCACTACTTAAGACCTGTGACTTTCGTTGCAACCTGCGTTG 2295













QY	1733	ACTACTGCTGCTCCCTGAATCCCAATCACTGTCACATCAGCGCCTTACGCTGCCCCGACC	1793
DB	501	CCTGCTCTCCGCCCTCTGCTCTGCTGCCCTCTGCTCTGCTGCCCTCTGCTGCCCT	442
QY	1793	GCTGCCCCCAGCGCTGCCGCCAGCGCCCTACTACTGTGTGCCAGAAATCCCACT	1852
DB	441	CTGCGCCCTCTGCCCTCTGCTGCTCTGCCCCCTCTGCTGCTGCCCTCTCTGCCCT	382
QY	1853	ACTGTCAATCACCACCTACTGTGTGCCCCCACTACCGCAGCACCTGCCCCCAAGACACA	1912
DB	381	CCTGCCCTCTCTGCTCTGCCCTCTGCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCT	322
QY	1913	GTACTGTACACCCACTGTGTGCCCTCTACTACCGCAGCACCTGTCTCCCAACACCACTG	1972
DB	321	CCTCTGCTCTCTGCCCTCTGCTGCCCTCTGCTGCTGCTGCCCTCTGCTGCTGCTGCC	262
QY	1973	ACTGCACACCCAGCGCCCTACTACCGCAGCACCTGCCCCCAACACCACTGCTGCT	2032
DB	261	CTGCTGCCCTCTGCCCTCTGCTGCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCT	202
QY	2033	GTACACCCCACTGCTGCCCCCACTGCAGCTGCCCTTCCCTGCGCACCTGCAACCAAGACC	2092
DB	201	GCCCCCTCTCTGCTCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCT	142
QY	2093	ACAGTGTCCCGGTACTACACACGAGCGCAGCTACACACCTGAAGATGATGACATC	2152
DB	141	CCTCTGCTCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCTCTGCTGCCCT	82
QY	2153	GACCCCTCTGCC	2166
DB	81	CCTCTCTCTGCTCC	68

RESULT 14

US-08-733-369A-112/c

; Sequence 112, Application US/08733369A

; GENERAL INFORMATION:

; APPLICANT: Masucci, Maria G.

; TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES CONFERRING

; TITLE OF INVENTION: INVISIBILITY TO THE IMMUNE SYSTEM.

; NUMBER OF SEQUENCES: 123

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kathleen M. Williams, Banner & Witcoff, Ltd.

; STREET: One Financial Center

; CITY: Boston

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02111

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Wordperfect 6.1a

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/733,369A

; FILING DATE: 17 October, 1996

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/522,995

; FILING DATE: 01-SEP-1995

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: US 08/529,190

; FILING DATE: 15-SEP-1995

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: SE 95013249

; FILING DATE: 10-APR-1995

; PRIORITY APPLICATION DATA:

; APPLICATION NUMBER: PCT/GB96/00876

; FILING DATE: 10-APR-1996

; ATTORNEY/AGENT INFORMATION:

; NAME: Williams, Kathleen M.

; REGISTRATION NUMBER: 34,380

; REFERENCE/DOCKET NUMBER: 95-1391-D

; TELECOMMUNICATION INFORMATION:



```
QY 2033 GTACCAACCCACTGCTGCCCCCTGCTGAGCTCCCGCTACCGTGGCAGCTGCACCCACACACC 2092
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 872  GCCCTTCCTCCTGCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCCCT 813
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2093 ACAGTGTCCCGGTAACTAAACCAACGCGCAGCTACCCACACCTGAAGATGATGACATC 2152
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 812  CCTCTGTCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCCCTCCTGCC 753
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 2153 GACCCCTCTCTCC 2166
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 752  CCTCTCTCTCTCC 739
Db      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: July 18, 1999, 02:09:55  
Job time: 3:09 sec

**THIS PAGE BLANK (USPTO)**



[illegible]

```

; SEQUENCE CHARACTERISTICS
;     LENGTH: 10596 bases
;     TYPE: NUCLEIC ACID
;     STRANDEDNESS: single
;     TOPOLOGY: linear
;
US-07-885-971-15

```

Query Match	5.6%;	Score 158.8;	DB 1;	Length 10596;
Best Local Similarity	53.3%;	Pred. No. 1.1e-24;		
Matches 359;	Conservative 0;	Mismatches 312;	Indels 3;	Gaps 1;

OY	1496	CCGACACTGTCTCACAAGGTGGCCCCAAGGGTGGCCCCAAGGGTGCCCAATGTGCACA	1555
Db	2946	CCTCACACTCTGTGTCTGTGCCCTCTGTCTGTCTGTGCCCTCTGTCTGTCTGTGCCCTCTGT	2887
OY	1556	CCCTTCACCGTGTGTCACCCTGTGCACGGCACCGCACCGCACTGCAGCCCCAGTCCACTACA	1615
Db	2886	GCCCCCTCTGTGTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGT	2827
OY	1616	ACCGCAATTCTTACTTCGGGCCCCCAGCCGCTGCCGCCACCGCACGAGTCTACTACTGTGCC	1675
Db	2826	CCCTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCC	2767
OY	1676	--CCTAAATCCCACACACTGTGCACAGTAACACACACTGCTGCTGCCAACGGACGCTCT	1732
Db	2766	CCCTCTGCCCTCTGTGTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCT	2707
OY	1733	ACTACTGCTGTCTCCCTGAATCCCATCTACTGTATCATACGAGCCTACCGGTCGCCCCAC	1792
Db	2706	CCCTGCTCTGTGCCCTCTGTGTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGTCTGTGCCCT	2647
OY	1793	GCTGCCCCCAACGCTGTGCCCCACCGACGCCCTACTACTGTCTGTCCAGAAAATCCAACT	1852
Db	2646	CCCTGCCCTCTGTGCCCTCTGTCTGTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCT	2587
OY	1853	ACTGTCAATATACACACTACTGTGTGCCCCCACTACCGAGACACTGGCCCCCAACACACA	1912
Db	2586	CCCTGCCCTCTGTGTCTGTGCCCTCTGTGTCTGTGTCTGTGCCCTCTGTGCCCTCTGTGCCCT	2527
OY	1913	GTCACGTATACACACACTGTGTGCCCTACTACCGAGACACTGTCCCAACACACACTGTG	1972
Db	2526	CCCTGTGTCTGTGTGCCCTCTGTGCCCTCTGTGTCTGTGTCTGTGTGCCCTCTGTGTCTGTGCC	2467
OY	1973	ACTGCACACCCACACCGAGCCCTACTATACCGAGACACTGTGCCCCCAACACACACTGTACT	2032
Db	2466	CCCTGTGTGCCCTCTGTGCCCTCTGTGTCTGTGTCTGTGTGCCCTCTGTGTCTGTGTGCCCT	2407
OY	2033	GTACACACCACTGTGTGCCCTCACTGTGACGTCCGCCCTACCGGTGCACCTGCACCCAAACCC	2092
Db	2406	GCCCCCTCTGTGTCTGTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGTCTGTGTGCCCT	2347
OY	2093	ACAGCTGCCCGGATACATAACACGAGCACGACATACACACGCGAAGATGATACATC	2152
Db	2346	CCCTGTGTCTGTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGCCCTCTGTGTCTGTGTGCC	2287
OY	2153	GACCCCTCTGCC	2166
Db	2286	CCCTCTGTGTGCC	2273
RESULT 3 US-08-087-783A-15/c ; Sequence 15, Application US/08087783A ; Patent No. 5547856 ; GENERAL INFORMATION: ; APPLICANT: Godowski, Paul J., Lokker, Nathalie A., Mark, Melanie R. ; TITLE OF INVENTION: HEPATOCTYE GROWTH FACTOR VARIANTS ; NUMBER OF SEQUENCES: 22 ; CORRESPONDENCE ADDRESS: ; ADDRESSEE: Genentech, Inc. ; STREET: 460 Point San Bruno Blvd ; CITY: South San Francisco ;			

STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 3.5 inch, 1.44 Mb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Winpatin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,783A  
FILING DATE: 13-JUL-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P0755779P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5416  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-087-783A-15

Query Match	5.6%	Score 158.8	DB 1	Length 10596
Best Local Similarity	53.3%	Pred. No. 1.1e-24		
Matches 359	Conservative 0	Mismatches 312	Indels 3	Gaps 1
QY 1496	CCGACAGCTGCTCCACACGCTGCCCCAACAGGCTGCGCCCAACAGCGTACGCCCCAAGCTAGCGCGCA	1555		
Db 2946	CTCTCACTCTCTGCTCTCTGCCCCCTCTGCTCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2887		
QY 1556	CCCTCCACCGTGTGCCACTGTGCAAGGCGACCGGCACTGTGACGAGCCCGCACTGCTCCACTTACA	1615		
Db 2886	GCCCCTCTCTGTCGCGCCCTCTCTGCT	2827		
QY 1616	ACCGCAATTCCTACTCTCGGGCCCCACCGCTGCCCCCAGCGAGCTCTACTACTGCTGCTGC	1675		
Db 2826	CTCTGCTCTCTGCGCCCT	2767		
QY 1676	---CCTGAATCCCAACACACGTGTACAGTACCACTACTGCTGTCTTCCACGCGAGCCCT	1732		
Db 2766	CTCTCTGCCCCCTCTGCTCTCTGCCCCCTCTCTGCCCCCTCTGCTCTCTGCCCCCTCTGCCCCCT	2707		
QY 1733	ACTACTGCTGTGCCGAATCCCAATCCCAATCACTGTCATGACATAGAGGCGCTACAGCGTCCCCCAGC	1792		
Db 2706	CTCTGCTCTGCCCCCTCTGCTCTGCCCCCTCTCTGCTCTGCCCCCTCTGCTCTGCCCCCTCTGCCCCCT	2647		
QY 1793	GCTGCCCCCAGCGTGCCTCCACCGCAGCCCTCTACTACTGCTGTGCCAAGAAATCCCAACT	1852		
Db 2646	CTCTGCCCCCTCTGCCCCCT	2587		
QY 1853	ACTGTCACTACCAACCACTACTGCTGCCCCCAGCTACCGCAGCAGCACTGCCCCCAACACACA	1912		
Db 2586	CTGTGCCCCCTCTGCTCTGCCCCCTCTCTGCTCTGCCCCCTCTGCTCTGCCCCCTCTGCCCCCT	2527		
QY 1913	GTCACGTACCAACCACTGCTGCCCCCTTACTTACCGAGACCTGCTTCCCAACACACAGTG	1972		
Db 2526	CTCTGCTCTCTGCTCTGCCCCCT	2467		
QY 1973	ACTGACCAACCAACCGAGCCCTCTACTACCGAGACACTGCCCCCAACACACAGTCACT	2032		

[illegible]

RESULT 4  
 US-08-194-088B-15/c  
 : Sequence 15 Application US/08194088B  
 : Patent No. 5580963  
 : GENERAL INFORMATION:  
 : APPLICANT: Godowsky I, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
 : TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS  
 : NUMBER OF SEQUENCES: 21  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Genentech, Inc.  
 : STREET: 460 Point San Bruno Blvd  
 : CITY: South San Francisco  
 : STATE: California  
 : COUNTRY: USA  
 : ZIP: 94080  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: patin (Genentech)  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/194,088B  
 : FILING DATE: 09-FEB-1994  
 : CLASSIFICATION: 530  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: 07/884811  
 : FILING DATE: 18-MAY-1992  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Gallegos, R. Thomas  
 : REGISTRATION NUMBER: 32,692  
 : REFERENCE/DOCKET NUMBER: 755D1  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 415/225-2614  
 : TELEFAX: 415/952-9881  
 : TELEX: 910/371-7168  
 : INFORMATION FOR SEO ID NO: 15:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 10596 bases  
 : TYPE: nucleic acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : US-08-194-088B-15

Query March	5.6%;	Score 158.8;	DB 1;	Length 10596;
Best Local Similarity	53.3%;	Pred. No. 1,1e-24;		
Matches 359;	Conservative	0;	Mismatches 312;	Indels 3;
			Gaps	1;

  

OY	1496	CCGACAGCTGCTCCACGAGGTGCCCCAACGGGTGCCCCACGCTGCCCCCACTGCGCGCA	1555
Db	2946	CCTCACCTCTCTGTGCTCTGCCCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT	2887
OY	1556	CCCTCCACACCGTGTCCACACTGCACAGCGCACCGCCCACTGTGAGGCCCGATCCCACTACA	1615
Db	2886	GCCCCCTCTGCTCTCTGCCCCCTCTCTGCCCCCTCTGCTCTGCCCCCTCTCTCTCTCTCT	2827
OY	1616	ACCGCAATTCCTACTCTCGGGCCCCACCGCTGCCCCCAACCGAGCTCCTACTACTGTGGCC	1675



Db	2836	CCTGCCCCCTCCTGGCCCTTCTCCAGTCGACGCCCTCCGACCCTCGCCCTCTCTCTGTGC	2767
Oy	1676	---CCTGAATPCCCACAACCACTGTGCAGAGTACACACTACTGTGTCTCCACGGCACCCCT	17322
Db	2766	CCTCTGGCCCTCTGTGCTGTGGCCCTCCGCCCCCTCCGCCCCCTCCCTCCGCCCTCTCTCCCT	2707
Oy	1733	ACTAAGTGTCCTCCCTGAATAATCCCATATACGTGCATATCAGCGGCTATACGGCTGCCGCCAC	17922
Db	2706	CCTGTCTCTGGCCCTTCTCTGCTCTTGCCCCCTCTGTCTGTCTGTCCCTCTGTCTGTGCCCT	26472
Oy	1793	GCTGCCCCCACAGCTGGCCCCACCGACGACCCCTACTACTGTCTGTCCAGAAATCCCACT	18552
Db	2646	CCTGGCCCTCTGTGGCCCTCCGCTCGTCTCTGTGCCCCCTCTGTCTGTGCCCTCTGTGCCCT	2587
Oy	1853	ACTGTCAATCACCACTACTGTGTGCCCCACTACCGCAGACCTGGCCCCAACAGACA	19122
Db	2586	CCTGCCCCCTCTGTCTGTGGCCCTCCCTCCGCTCGTCTGTGCCCCCTCCCTCTGTGCCCT	2527
Oy	1913	GTCACTGTACCAACCACTGTGTGCCCTCTACTACCGCAGACCTGTCTCCAACACCACTG	19722
Db	2526	CCTCTGTCTCTGTGCCCTCTGTGCCCTCCCTCCGCTCGTCTGTGCCCTCTGTCTGTGCCCT	2467
Oy	1973	ACTGCACCAACCCACCGAGGCCCTACTACCGCAGACCTGGCCCCAAGCACCACTACT	20322
Db	2466	CCTCTGTGCCCTCTGTGCCCTCTCTGTCTGTCTGTGCCCTCTGTGCCCTCTGTCTGTCT	2407
Oy	2033	GTACCAACCACTGTGTGCCCCACTGTGACGTCCCTCTACACGTGTGACCTGTGACCAACAC	20922
Db	2406	GCCCTCTCTCTGTGTCTGTGGCCCTCCGCCCCCTGTGCCCCCTCTGTCTGTGCCCT	2347
Oy	2093	ACAGCTGCCCGGGTAACTATACACCGGACCGACGCTAACACACTGTGAAGATGATGATC	21522
Db	2346	CCTCTGTCTCTGTGCCCTCTGTGCCCTCCGCCCCCTCTGTGCCCTCTGTCTGTGCCCT	2287
Oy	2153	GACCCCTCTGCC	2166
Db	2286	CCTCTCTCTGTCTC	2273

RESULT 5  
 US-08-194-087-15/c  
 : Sequence 15 Application US/08194087  
 : Patent No. 5879910  
 :  
 : GENERAL INFORMATION:  
 : APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
 : TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR PROTEASE DOMAIN VARIANTS  
 : NUMBER OF SEQUENCES: 21  
 :  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Genentech, Inc.  
 : STREET: 460 Point San Bruno Blvd  
 : CITY: South San Francisco  
 : STATE: California  
 : COUNTRY: USA  
 : ZIP: 94080  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: patin (Genentech)  
 :  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/194,087  
 : FILING DATE: 18-May-1992  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER:  
 : FILING DATE:  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Dregert, Ginger R.  
 : REGISTRATION NUMBER: 33,055  
 : REFERENCE/DOCKET NUMBER: 779  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 415/225-3216  
 : TELEFAX: 415/952-9881  
 :

```

: TELEX: 910/371-7168
: INFORMATION FOR SEQ ID NO: 15:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 10596 bases
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
:
US-08-194-087-15

```

Query Match	5.6%	Score 158.8	DB 4	Length 10596
Best Local Similarity	53.3%	Pred. No. 1.1e-24		
Matches 359	Conservative 0	Mismatches 312	Indels 3	Gaps 1

[illegible]

RESULT 6  
PCT-US93-04648-15/C  
Sequence 15, Application PC/TUS9304648  
GENERAL INFORMATION:  
APPLICANT: Genentech, Inc., Godowski, Paul J., Lokker, Natalie A., Mark, Melanhi  
TITLE OF INVENTION: HEPATOCYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd



Db 2120 CAGCGGCTCAGCTCTCGAGGGGCTCCAGACACACACCTTACCTGATACACACAGCCG 2061  
QY 1542 CCCCACTGCGGACCTCCAGCGTGGTCCGACCTGACACGCGACCGGACCTGCGAGCCG 1601  
Db 2060 CTCGGATCTGACAGAGCTGACAGACAGACACCTCTGCTCGGAGCGGTATCCAG 2001  
QY 1602 CAGTCCACCTACACACGCAATTTCTACTCCGCGCCCAACCGTGGCCCGACGAGCTC 1661  
Db 2000 CTTGACACACGAGAGCTTCAGAGCTGCTGCTGACCTGAGAGGCTCCAGCTCCAGCAG 1941  
QY 1662 CTACTACTGCTGCGCTGATATCCCAACAGCTGTACAGTACCACTGCTGCTCCCA 1721  
Db 1940 CAGCGGCTCTGACAGACACCTCTACTGTAAACCCGACGCGCTCCAGCTCTGCTCCAG 1881  
QY 1722 CCGAGCGGCTCTACTGCTGTGCTCCGAATGCCAATCTACTGTCAATGAGGCGCTACCG 1781  
Db 1880 CAGCTGCGCGAGACAGAGTCTTGTGCTCGGACCGATATCCACCTGACACCGGTAGCTC 1821  
QY 1782 CTGCCCCACCGCTGCGCCCGACCGCTGCGCCACCGACCGCCCTACTGCTGTGCTCCAG 1841  
Db 1820 CAGCTGCTGCTCTGACAGCTGAGGCGCTCCAGCTCCAGACCTGACAGAGCTCCAGCTC 1761  
QY 1842 AATGCCAATCTACTGTACATACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1901  
Db 1760 CAGACACACAGACCTCTACAGTACAGACAGACCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1701  
QY 1902 CCAACACACAGCTACTGTACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1952  
Db 1700 CACGAGCTCTCTGCTCGGACACCTATACCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1641  
QY 1953 CTGCTCCCAACACAGAGTACAGACAGACAGACCGACCGCCCTACTGCTGCTGCTGCTGCT 2012  
Db 1640 CTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1581  
QY 2013 CCCCCAACACAGTACTGTACAGACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2072  
Db 1580 CCGCT 1521  
QY 2073 TCGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2132  
Db 1520 CACCTTGAACACAGTACGCTTCCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1461  
QY 2133 CACCT 2137  
Db 1460 CACCT 1456

RESULT 8  
US-08-458-298-1/C  
; Sequence 1, Application US/08458298  
; Patent No. 5756677  
; GENERAL INFORMATION:  
; APPLICANT: Lewis, Randolph V.  
; APPLICANT: Colgin, Mark  
; TITLE OF INVENTION: CDNA Encoding Minor Ampullate Spider  
; NUMBER OF SEQUENCES: 56  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Birch, Stewart, Kolasch & Birch  
; STREET: P.O. Box 747  
; CITY: Falls Church  
; STATE: Virginia  
; COUNTRY: USA  
; ZIP: 22040-3487  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/458,298  
; FILING DATE: 02-JUN-1995

CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/209,747  
FILING DATE: 14-MAR-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8000  
TELEFAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHEICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Nephila clavipes  
TISSUE TYPE: minor ampullate gland  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 183..2675  
OTHER INFORMATION: /product= "N. clavipes minor  
OTHER INFORMATION: ampullate silk protein"  
US-08-458-298-1

Query Match 4.3%; Score 120.2; DB 3; Length 2793;  
Best Local Similarity 50.1%; Pred. No. 8.9e-17;  
Matches 333; Conservative 0; Mismatches 323; Indels 9; Gaps 1;

QY 1482 CATGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1541  
Db 2120 CAGCGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2061  
QY 1542 CCGCACTGCGCGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1601  
Db 2060 CTCGGATCTGACAGAGCTGACAGAGCTGACAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2001  
QY 1602 CAGTCCACCTACACACGCAATTTCTACTCCGCGCCCAACCGTGGCCCGACGAGCTC 1661  
Db 2000 CTTGACACACGAGAGCTTCAGAGCTGCTGCTGACCTGAGAGGCTCCAGCTCCAGCTC 1941  
QY 1662 CTACTACTGCTGCGCTGATATCCCAACAGCTGTACAGTACCACTGCTGCTCCCA 1721  
Db 1940 CAGCGGCTCTGACAGACACCTCTACTGTAAACCCGACGCGCTCCAGCTCTGCTCCAG 1881  
QY 1722 CCGAGCGGCTCTACTGCTGTGCTCCGAATGCCAATCTACTGTCAATGAGGCGCTACCG 1781  
Db 1880 CAGCTGCGCGAGACAGAGTCTTGTGCTCGGACCGATATCCACCTGACACCGGTAGCTC 1821  
QY 1782 CTGCCCCACCGCTGCGCCCGACCGCTGCGCCACCGACCGCCCTACTGCTGTGCTCCAG 1841  
Db 1820 CAGCTGCTGCTCTGACAGCTGAGGCGCTCCAGCTCCAGACCTGACAGAGCTCCAGCTC 1761  
QY 1842 AATGCCAATCTACTGTACATACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1901  
Db 1760 CAGACACACAGACCTCTACAGTACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1701  
QY 1902 CCAACACACAGTACTGTACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1952  
Db 1700 CACGAGCTCTCTGCTCGGACACCTATACCGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1641  
QY 1953 CTGCTCCCAACACAGAGTACAGACAGACAGACCGACCGCCCTACTGCTGCTGCTGCTGCT 2012  
Db 1640 CTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1581  
QY 2013 CCCCCAACACAGTACTGTACAGACAGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2072  
Db 1580 CCGCT 1521



HYPOTHETICAL: NO  
ANTI-SENSE: NO  
PCT-US93-03993-1

Query Match 3.9%; Score 110.8; DB 5; Length 6530;  
Best Local Similarity 48.4%; Pred. No. 1.2e-14;  
Matches 307; Conservative 0; Mismatches 327; Indels 0; Gaps 0;

1500 CAGCTGCTCCACAGGCTGCGCCACAGCGCTGCGCCCAACAGGCTGCGCCCACTGCGCGACCGCT 1559  
1560 CCACGCTGCTCCACAGGCTGCGCCACAGCGCTGCGCCCAACAGGCTGCGCCCACTGCGCGACCG 1619  
1560 CCACGCTGCTCCACAGGCTGCGCCACAGCGCTGCGCCCAACAGGCTGCGCCCACTGCGCGACCG 1619  
3430 CCGCGCGCGCGCGAGCTGATCCACCGCAGCTGATCCACCGCAGCTGATCCACCGCAGCTGATCC 3371  
1620 CAATTCCTACTCG 1679  
3370 CCGCTGCTTCCACCG 3311  
1680 AATCCCAACAGCTGTCACAGTACACAGCTGTCACAGCTGTCACAGCTGTCACAGCTGTCACAG 1739  
3310 GAGTACTGCTGCGCTGCG 3251  
1740 CTGTCCCTGAAATCCCAATCTGTCACATGTCACATGTCACATGTCACATGTCACATGTCACAT 1799  
3250 CCGCGCGAGTACAGAGCG 3191  
1800 CCACGCTGCTCCACAGGCTGCGCCACAGCGCTGCGCCCAACAGGCTGCGCCCACTGCGCGAC 1859  
3190 CCGCGCGAGAGCG 3131  
1860 CATCAACAGCTGTCACAGTACACAGCTGTCACAGCTGTCACAGCTGTCACAGCTGTCACAG 1919  
3130 AGCG 3071  
1920 TACACCGCTGCTGCG 1979  
3070 CCGCGCGAGTACAGAGCG 3011  
1980 CACCGCTGCTCCACAGGCTGCGCCACAGCGCTGCGCCCAACAGGCTGCGCCCACTGCGCGAC 2039  
3010 CCGCGCGAGAGTACAGAGCG 2951  
2040 CCACGCTGCTCCACAGGCTGCGCCACAGCGCTGCGCCCAACAGGCTGCGCCCACTGCGCGAC 2099  
2950 CCGCGAGTACAGAGCG 2891  
2100 CCGCGTACTACAGAGCG 2133  
2890 CCGCGAGTACAGAGCG 2857

RESULT 11  
US-08-056-200-93/c  
Sequence 93, Application us/08056200  
Patent No. 5616500

GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
TITLE OF INVENTION: Methods of Using Same  
NUMBER OF SEQUENCES: 117  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor  
CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.

ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054,001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEQ ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 951 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
NAME/KEY: Intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93

Query Match 3.7%; Score 105; DB 1; Length 951;  
Best Local Similarity 49.3%; Pred. No. 2.4e-13;  
Matches 310; Conservative 0; Mismatches 310; Indels 9; Gaps 1;

1461 GCGATTGACAAACAGTGCATGACTGTCGCGCGAGACAGCTGTCGCGCGAGCTGCGCG 1520  
3810 GCGCTTACAGCAATGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3751  
1521 CAAGCGTGGCGCGAGCTGCGCGCGAGCTGCGCGCGAGCTGCGCGCGAGCTGCGCGAG 1580  
3750 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3691  
1581 GCGCACCGCGCAACAGTGCATGACTGTCGCGCGAGACAGCTGTCGCGCGAGCTGCGCGAG 1640  
3690 CCGTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3631  
1641 CCGTGGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1700  
3630 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3571  
1701 TACACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1760  
3570 CTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3511  
1761 CTGTACATCAAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1820  
3510 GCGCTTACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3451  
1821 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1874  
3450 CCGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3391  
1874 --GCTGCGCGCGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1931  
3390 CAGCT 3331



STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Gossypium barbadense  
STRAIN: Sea Island  
IMMEDIATE SOURCE:  
LIBRARY: EMBL-SI  
CLONE: SIH6  
US-08-298-687A-25

Query Match 3.6%; Score 101; DB 1; Length 1985;  
Best Local Similarity 51.4%; Pred. No. 8.6e-13;  
Matches 233; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 1581 CGCCACCGGCACTGACGCGCCAGTCCACATCAACCGCAATTCCTACTCGGCCCCCA 1640  
DB 424 CGGCACCGCACACCGCGTTTCTACTCTCTCCGAGACTCAAGCAGCGCTACACCAA 483  
QY 1641 CCGCTGCCCGCCAGCGAGCTCTACTACTGTGCGCCCTGAATCCCAACGACTGTACAG 1700  
DB 484 CCGCCACTCGCGCAGCGGTTTCTACTCTCTCCACATTCATCACGCCGCCAGTGACAG 543  
QY 1701 TACCACTACTGTCTCTCCACCGCAGCGCCCTACTACTGTCTCTGAAATCCCAATCA 1760  
DB 544 CTCTTCACACCGCCAGTTTCACTCTCCACCGCAGTTCTCTCTCTGCACTCCACAC 603  
QY 1761 CTGTACATCAGCGCGCTTACCGTCCCGCAGCGCTGCCCGCCAGCGCCAGCGCAG 1820  
DB 604 CTGCTCTCTCTCTCTCTGCACTCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTC 663  
QY 1821 CCCCTACTACTGTCTCTCCAGGAATCCCACTACTGTCTCTCTCTCTCTCTCTCTCT 1880  
DB 664 CAGCTTCT 723  
QY 1881 CCACCTACCGCAGCGCTGCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1940  
DB 724 CTGCTACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 783  
QY 1941 CTACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2000  
DB 784 CAGTACAAACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 843  
QY 2001 CCGCAGCAGCTGCGCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2033  
DB 844 TCGGGGCTGTACGCGCAGGTCCAGTGAAGTCTG 876

RESULT 14  
US-08-298-829-25  
Sequence 25, Application US/08298829  
Patent No. 5620882  
GENERAL INFORMATION:  
APPLICANT: John, Maliyakal E.  
TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON  
TITLE OF INVENTION: PLANTS FOR ALTERED FIBER  
NUMBER OF SEQUENCES: 33  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
STREET: P.O. Box 2113, First Wisconsin Plaza  
CITY: Madison  
STATE: Wisconsin  
COUNTRY: USA  
ZIP: 53701  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Microsoft Word  
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/298, 829  
FILING DATE: 19-OCT-1994  
CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/885, 970  
FILING DATE: 18-MAY-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/617, 239  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/253, 243  
FILING DATE: 04-OCT-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27,386  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 283-2478  
TELEFAX: (608) 251-5139  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1985 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Gossypium barbadense  
STRAIN: Sea Island  
IMMEDIATE SOURCE:  
LIBRARY: EMBL-SI  
CLONE: SIH6  
US-08-298-829-25

Query Match 3.6%; Score 101; DB 1; Length 1985;  
Best Local Similarity 51.4%; Pred. No. 8.6e-13;  
Matches 233; Conservative 0; Mismatches 220; Indels 0; Gaps 0;

QY 1581 CGCCACCGGCACTGACGCGCCAGTCCACATCAACCGCAATTCCTACTCGGCCCCCA 1640  
DB 424 CGGCACCGCACACCGCGTTTCTACTCTCTCCGAGACTCAAGCAGCGCTACACCAA 483  
QY 1641 CCGCTGCCCGCCAGCGAGCTCTACTACTGTGCGCCCTGAATCCCAACGACTGTACAG 1700  
DB 484 CCGCCACTCGCGCAGCGGTTTCTACTCTCTCTCCACATTCATCACGCCGCCAGTGACAG 543  
QY 1701 TACCACTACTGTCTCTCCACCGCAGCGCCCTACTACTGTCTCTGAAATCCCAATCA 1760  
DB 544 CTCTTCACACCGCCAGTTTCACTCTCCACCGCAGTTCTCTCTCTGCACTCCACAC 603  
QY 1761 CTGTACATCAGCGCGCTTACCGTCCCGCAGCGCTGCCCGCCAGCGCCAGCGCAG 1820  
DB 604 CTGCTCTCTCTCTCTCTGCACTCTCCACCTCTCTCTCTCTCTCTCTCTCTCTCTC 663  
QY 1821 CCCCTACTACTGTCTCTCCAGGAATCCCACTACTGTCTCTCTCTCTCTCTCTCTCT 1880  
DB 664 CAGCTTCT 723  
QY 1881 CCACCTACCGCAGCGCTGCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1940  
DB 724 CTGCTACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 783  
QY 1941 CTACCGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2000  
DB 784 CAGTACAAACACCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 843  
QY 2001 CCGCAGCAGCTGCGCCCGCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 2033  
DB 844 TCGGGGCTGTACGCGCAGGTCCAGTGAAGTCTG 876

```

RESULT 15
US-08-217-327-3
: Sequence 3, Application US/08217327
: Patent No. 5474925
: GENERAL INFORMATION:
: APPLICANT: John, Maliyakal E
: APPLICANT: Barton, Kenneth A
: TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Charles and Brady
: STREET: P.O. Box 2113
: CITY: Madison
: STATE: WI
: COUNTRY: USA
: ZIP: 53701-2113
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/217,327
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 07/812,233
: FILING DATE: 19-DEC-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Seay, Nicholas J
: REGISTRATION NUMBER: 27,386
: REFERENCE/DOCKET NUMBER: 1122990831
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 608-251-5000
: TELEFAX: 608-251-9166
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 913 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Gossypium hirsutum
: STRAIN: Coker 312
: TISSUE TYPE: Fiber cells
: IMMEDIATE SOURCE:
: CLONE: H6
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 72..716
: FEATURE:
: NAME/KEY: sig_peptide
: LOCATION: 72..147
: US-08-217-327-3

Query Match 3.5%, Score 100; DB 1; Length 913;
Best Local Similarity 51.8%; Pred. No. 1e-12;
Matches 226; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1581 GCGCACCGGCACTGACCCCAAGTCCACACCAACCGCAATTCCTACTCGGGCCCA 1640
DB 175 CGACACCGACACCGGCTTCTACTCTCTCGACGACTCAAGCACCCTTACACCAA 234
QY 1641 CCGTGCCCCACCGACCTCTACTACTGCTGCGCTGAATCCCAACCACTGTACAG 1700
DB 235 CCGCACCTCGGACCGGTTTACTCTCTCCACCTTACATCAGCGCCCGAGTACAG 294
QY 1701 TACACCTACTGCTGCTCCACCGACGCGCTACTACTGCTGCTCCGAATCCCAATCA 1760

```

```

DB 295 CTCTCACCACCCGAGTTTCAACTCTCTCCACCCAGCTTCTCTCTGCAACTCCACAC 354
QY 1761 CTGTACATCAGCGCGCTACCGCTGCGCCCAACCGCTGCGCCACCGCTGCGCCAG 1820
DB 355 CTGCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 414
QY 1821 CCGCTACTACTGCTGTCGCCAAGAAATCCCACTACTGTGTCATCAGCACTACTGTC 1880
DB 415 CAGCTTCTCAGCTTCCGCGCACTGCACTGCACTGCACTGCACTGCACTGCACTGCA 474
QY 1881 CCATACCGGACGACCTGCGCCCAACACACAGATGACTGACACCGACCGACCGCTACT 1940
DB 475 CTGCTACCCGACACACAGCTCTGCTTCTCTCTGACGACGCTCTCTCTCTCTCTCT 534
QY 1941 CTACCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2000
DB 535 CAGTACAAACACCAATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 594
QY 2001 CCGCAGCACCTGCCCC 2016
DB 595 TCGGGGCTGCTACGCG 610

```

Search completed: July 18, 1999, 07:54:43  
Job time: 384 sec



**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 11:07:49 ; Search time 133.54 Seconds  
(without alignments)  
3974.226 Million cell updates/sec

Title: US-09-294-663-2  
Perfect score: 2821  
Sequence: 1 GAAAGAGATACACGCGAC.....TAAAAAAAAAAAAAAAAAAAA 2821  
Scoring table: IDENTITY\_NUC

Searched: 240622 seqs, 94065609 residues  
Database : N.Geneseq\_34.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	158.8	5.6	10596	1 Q51731	Plasmid pcISEBON f
2	158.8	5.6	10596	1 T40348	Plasmid pcISEBON f
3	158.8	5.6	9600	1 V21683	Vector plasmid PCM
4	158.8	5.6	799	1 V55831	Nucleotide sequenc
5	151.6	5.4	795	1 V55830	FLGA insert stabl
6	137.8	4.9	5163	1 V20700	Cryptosporidium pa
7	136.2	4.8	5118	1 V20701	Cryptosporidium pa
8	120.2	4.3	2744	1 O59470	Loricrin gene. Con
9	110.4	3.9	6530	1 O51557	MisP1-containing p
10	110.4	3.9	40875	1 T80043	Insert from cosmid
11	101	3.6	1984	1 T13030	Cotton fibre-speci
12	101	3.6	1985	1 T30250	Cotton fibre-speci
13	101	3.6	1985	1 T70036	Cotton fibre clone
14	100.8	3.6	5008	1 O27223	Cotton H6 gene and
15	100	3.5	913	1 T07199	MSF precursor. New
16	100	3.5	913	1 T13034	Cotton fiber-speci
17	100	3.5	913	1 T30253	Cotton fibre-speci
18	100	3.5	913	1 T62610	Cotton fibre cell-
19	100	3.5	913	1 T70041	Cotton fibre speci
20	98.8	3.5	502	1 O06483	Intestinal mucin c
21	97.4	3.5	1368	1 N91235	DNA sequence of pr
22	95	3.4	1107	1 V38816	DNA encoding a dur
23	93	3.3	860	1 O06486	Intestinal mucin c
24	92.4	3.3	3331	1 V22683	New DNA sequence i
25	92.4	3.3	2214	1 V22682	New DNA sequence i
26	91.4	3.2	859	1 O12220	SMUC-53 intestinal
27	90.8	3.2	502	1 O12217	SMUC-40 intestinal
28	90.4	3.2	44377	1 T78508	Platenolide synth
29	90.4	3.2	13987	1 T80415	Hybrid smg/tylg O
30	90.4	3.2	44377	1 T80414	Nephila clavipes s
31	89	3.2	2004	1 T85356	Genomic clone G1F
32	88.8	3.1	1505	1 O55750	Intestinal mucin c
33	86.6	3.1	836	1 O06484	Intestinal mucin c
34	86.6	3.1	836	1 O12218	SMUC-41 intestinal
35	85.6	3.0	2550	1 V52497	Streptococcus pneu
36	85.6	3.0	2000	1 N71065	Gene encoding plas
37	84.4	3.0	12001	1 O76213	HSV L/SF region. H
38	83	2.9	390	1 O21833	Randomising oligon
39	83	2.9	390	1 O36859	PCR primer for 5'
40	82.6	2.9	1702	1 V52500	Streptococcus pneu
41	82.4	2.9	32768	1 V52204	Streptococcus pneu
42	80.4	2.9	14704	1 O20685	PKS 741 Insert con
43	80.2	2.8	4483	1 V52496	Streptococcus pneu

ALIGNMENTS

C 44	80	2.8	43280	1 T80413
C 45	79.8	2.8	2338	1 V23249
Ty lactone synthase Nephila clavipes s				
ALIGNMENTS				
RESULT 1				
051731/C				
ID	051731 standard; DNA; 10596 BP.			
AC	051731:			
DT	31-MAY-1994 (first entry)			
DE	Plasmid pcISEBON for subcloning huGF variants.			
KW	Hepatocyte Growth Factor; HGF; variant; mutagenesis; in vitro mutagenesis;			
KW	proteolysis resistant; liver; malignancy; CMV-driven;			
KW	Cytomegalovirus; episomal expression plasmid; ss.			
OS	Synthetic.			
PH	Key			
FT	enhancer			
FT	1..611			
FT	Location/Qualifiers			
FT	/tag= a			
FT	/note= "CMV enhancer/promoter"			
FT	758..775			
FT	/tag= b			
FT	/label= SP6_promoter			
FT	845..849			
FT	/tag= c			
FT	/note= "SP6 RNA start"			
FT	902..966			
FT	/tag= d			
FT	/function= cloning_linker			
FT	967..1107			
FT	/tag= e			
FT	/note= "SV40 poly A"			
FT	1108..1531			
FT	/tag= f			
FT	/function= SV40_origin			
FT	1580..4189			
FT	/tag= g			
FT	/label= EBNA-1			
FT	4190..6374			
FT	/tag= h			
FT	/function= oriP			
FT	4295..4887			
FT	/tag= i			
FT	/note= "family of repeats"			
FT	5866..5978			
FT	/tag= j			
FT	/note= "dyad region"			
FT	6375..6457			
FT	/tag= k			
FT	/label= HSV_TK_terminator_3'-end			
FT	6975..7975			
FT	/tag= l			
FT	/phenotype= neomycin_resistance			
FT	/note= "Tns neomycin phosphotransferase gene"			
FT	7975..8112			
FT	/tag= m			
FT	/label= TK_promoter			
FT	8114..8594			
FT	/tag= n			
FT	/function= M13_ori			
FT	8595..10414			
FT	/tag= o			
FT	/label= delta_2a			
PN	W09323541-A.			
PD	25-NOV-1993.			
PF	17-MAY-1993; U04648.			
PR	18-MAY-1992; US-884811.			
PR	18-MAY-1992; US-885971.			
PA	(GETH ) GENENTECH INC.			
PI	Godowski PJ, Lokker NA, Mark MR;			
DR	WPI: 93-386573/48.			
Hepatocyte growth factor variants are resistant to proteolytic				



Dn	2526	CCTGTGTCGTCTGCACCCCTCCTGGACCCTCCCTCCAGTCCGATCGAGGCCCTCTGCTCTGCTGC	2467
Oy	1973	ACTGCAACACCCACGACGACCCCTTAACCGACAGACCTGCCCCAAACACAAGTCACT	20322
Dn	2466	CCCTCTGTCCTCTGCTGCACCCCTCCCTCCGATCGATCGCCCTCCAGCCCTCCCTCTGCTCT	24078
Oy	2033	GTAGCACCACTGCTGGCCCCACTGCAGCTCCCCAACGCTGGACACCTGACACCCAAACAC	20922
Dn	2406	GCCCCCTCTCTGCTGCTCCGCCCCTCCCTGCACCCCTCCGCCCCTCCCTCCCTGCTGCCCCCT	23477
Oy	2093	ACACCTGCCCCGGTAATAATACACAGCGCACGTCACACACACATGAAGATGATGATCATC	21522
Dn	2346	CCCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	22877
Oy	2153	GACCCCTCTCTCC	2166
Dn	2286	CCCTCTGCTGCTC	2273
RESULT	3		
V21683/c			
ID	V21683	standard; DNA; 9600 BP.	
AC	V21683;		
DT	17-AUG-1998	(first entry)	
DE	Vector plasmid pCMVkmTr-EPI.		
KW	Polynucleotide delivery; plasmid pCMVkmTr-EPI; vector;		
OS	Chimeric - Epstein-Barr virus.		
OS	Chimeric - Adeno-associated virus.		
OS	Chimeric - Cytomegalovirus.		
FH	Key	Location/Qualifiers	
FT	CDS	14..2594	
FT	/tag-a	a	
FT	/product=	"EBV nuclear antigen A"	
FT	misc.feature	2623..4559	
FT	/tag-b	b	
FT	/note=	"EBV origin of replication"	
FT	repeat_unit	4928..5104	
FT	/tag-c	c	
FT	/rpt_type=	INVERTED	
FT	/note=	"AAV inverted terminal repeat"	
FT	repeat_unit	7189..7355	
FT	/tag-d	d	
FT	/rpt_type=	INVERTED	
FT	/note=	"AAV inverted terminal repeat"	
FT	promoter	5112..6734	
FT	/tag-e	e	
FT	/note=	"CMV immediate-early enhancer/promoter"	
FT	terminator	6818..7050	
FT	/tag-f	f	
FT	/note=	"bovine growth hormone polyA sequence"	
Pn	M09806437-A2.		
Dn	19-FEB-1998.		
PF	13-AUG-1997:	U14465.	
PR	13-AUG-1996:	US-023867.	
PA	(CHIR ) CHIRON CORP.		
PI	Cohen F, Dubois-Stringfellow N, Dworki V, Innis MA,		
PI	Murphy JF, Tetsuo U, Zukermann R;		
DR	WPI: 98-159296/14.		
PT	Polycationic agents based on alpha-amino acids, able to complex		
PT	with nucleic acid - to facilitate its entry into cell, condense it		
PT	and protect it against serum degradation, particularly for use in		
PT	gene therapy		
PS	Disclosure; Page 77-80: 100pp: English.		
CC	This polynucleotide comprises the DNA sequence of vector plasmid		
CC	pCMVkmTr-EPI, which contains an Epstein-Barr virus (EBV) origin		
CC	of replication from plasmid pCEB4, a coding region for EBV nuclear		
CC	antigen A from pCEB4, a pair of inverted terminal repeats from		
CC	adeno-associated virus, a cytomegalovirus enhancer/promoter, a		
CC	bovine growth hormone polyA sequence, and a kanamycin resistance		
CC	selectable marker. Polynucleotides encoding polypeptides, such as		
CC	erythropoietin or leptin, and ribozymes and antisense		

[illegible]



QY 1403 AGTCTGCGCAACAGTCTGTAGTCTCTGTTACCGCTGTTGCCAAGTTTTCAGATGC 1462  
 DB 777 ATTCCTGCGAGCCCCGCGCTCTCACTGCTCTCGACCCGCGCTCACTCTGCTCTGC 718  
 QY 1463 GATTCGACACAGTGCATGACGTGCTGCGCGCGAGCTCTCTCAAGCGTGCGCCA 1522  
 DB 717 CCTCTGCTCTGCGCGCT 658  
 QY 1523 ACGGCTGCGCGCAACGCGCTGCGCGCACTGCGCGCACTGCGCGCACTGCGCGCA 1582  
 DB 657 CCGGCGCGCT 601  
 QY 1583 CCAACCGCACTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1642  
 DB 600 CCGCTGCT 541  
 QY 1643 GCTGCGCGCGCAACGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1702  
 DB 540 CCGGCGCGCT 481  
 QY 1703 CCACTGCTGCTGCT 1762  
 DB 480 GCTGCTGCGCGCT 421  
 QY 1763 GTCACATGAGCGCGCT 1822  
 DB 420 CCGCTGCT 361  
 QY 1823 CCGTCTGCTGCT 1882  
 DB 360 GCGCGCT 301  
 QY 1883 ACTAGCGAGCACTGCTAC 1941  
 DB 300 CCGGCGCGCT 241  
 QY 1942 TACCGGAGCACTGCT 2001  
 DB 240 TCGCTCTGCT 181  
 QY 2002 CCGAGCACTGCG 2061  
 DB 180 CCGCTCTGCGCGCT 121  
 QY 2062 TCGCGCGCT 2121  
 DB 120 TCGCGCGCT 61  
 QY 2122 ACCAGCTACCACTGAGATGATG 2147  
 DB 60 TCGCT 35  
 RESULT 6  
 V20700 standard: DNA: 5163 BP.  
 AC V20700:  
 DT 17-AUG-1998 (first entry)  
 DE Cryptosporidium parvum GP900 antigen open reading frame.  
 KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;  
 KW antibody; prophylaxis; treatment; inhibition; retardation;  
 OS detection; diagnosis; human; ds.  
 OS Cryptosporidium parvum.  
 FH key  
 PH Location/Qualifiers  
 FT 1..5163  
 FT CDS  
 FT misc\_feature  
 FT 524..1270  
 FT /tag= a  
 FT /product= GP900 antigen  
 FT /tag= b  
 FT /note= "region containing NINC mutations"  
 PN W09806430-A1.  
 PD 19-FEB-1998.  
 PF 11-AUG-1997; U14104.

PR 14-AUG-1996; US-700651.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Gut J, Leech J, Nelson RC, Petersen C;  
 DR WPI: 98-159290/14.  
 DR P-PSDB: W48299.  
 PT Anti-Cryptosporidium antibody - used to develop products for  
 PT detection, diagnosis, prophylaxis or treatment of cryptosporidium  
 PT infections  
 PS Claim 32: Pages 60-62; 89pp: English.  
 CC The sequence is that encoding the GP900 antigen which may be used  
 CC in the production of anti-cryptosporidium antibodies. These can be  
 CC used for the prophylaxis, treatment, inhibition or retardation of  
 CC a cryptosporidium infection in humans or in animals such as calves.  
 CC They can also be used for the detection and diagnosis of related  
 CC infections.  
 SO Sequence 5163 BP; 1875 A; 1137 C; 873 G; 1278 T;  
 Query Match 4.9%; Score 137.8; DB 1; Length 5163;  
 Best Local Similarity 51.0%; Pred. No. 1.7e-16;  
 Matches 325; Conservative 0; Mismatches 312; Indels 0; Gaps 0;  
 QY 1481 ACATGCACTGCTGCT 1540  
 DB 584 AC 643  
 QY 1541 GCG 1600  
 DB 644 ACTGAC 703  
 QY 1601 CCAGTCCAC 1660  
 DB 704 AC 763  
 QY 1661 CCTACTGCTGCTGCT 1720  
 DB 764 ACTGAC 823  
 QY 1721 ACG 1780  
 DB 824 ACCAC 883  
 QY 1781 GCTGCG 1840  
 DB 884 AC 943  
 QY 1841 GAATCCAC 1900  
 DB 944 AC 1003  
 QY 1901 CCCAC 1960  
 DB 1004 ACTGAC 1063  
 QY 1961 AAC 2020  
 DB 1064 ACCAC 1123  
 QY 2021 ACCACAGTCTGCTGCT 2080  
 DB 1124 AC 1183  
 QY 2081 GCAC 2117  
 DB 1184 CCAC 1220  
 RESULT 7  
 V20701 standard: DNA: 5318 BP.  
 AC V20701:  
 DT 17-AUG-1998 (first entry)  
 DE Cryptosporidium parvum GP900 antigen open reading frame and 3' region.

KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;  
KM antibody; prophylaxis; treatment; inhibition; retardation;  
KW detection; diagnosis; human; 3' region; ds.  
OS Cryptosporidium parvum.  
FH Key Location/Qualifiers  
FT 1. .5166  
FT /\*tag= a  
FT /product= GP900 antigen  
FT misc\_feature 524. .1270  
FT /\*tag= b  
FT /note= "region containing NINC mutations"  
PN W09806430-A1.  
PD 19-FEB-1998.  
PF 11-AUG-1997; U14104.  
PR 14-AUG-1996; US-700651.  
PA (REGC ) UNIV CALIFORNIA.  
PI Gut J. Leech J, Nelson RC, Petersen C;  
DR WPI; 98-159290/14.  
DR P-PSDB; W48299.  
PT Anti-Cryptosporidium antibody - used to develop products for  
PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium  
PS Claim 32: Pages 62-63; 89pp; English.  
CC The sequence is that encoding the GP900 antigen which may be used  
CC in the production of anti-Cryptosporidium antibodies. These can be  
CC used for the prophylaxis, treatment, inhibition or retardation of  
CC a Cryptosporidium infection in humans or in animals such as calves.  
CC They can also be used for the detection and diagnosis of related  
CC infections.  
SQ Sequence 5318 BP; 1944 A; 1153 C; 888 G; 1333 T;  
Query Match 4.8%; Score 136.2; DB 1; Length 5318;  
Best Local Similarity 50.9%; Pred. No. 3,3e-16;  
Matches 324; Conservative 0; Mismatches 313; Indels 0; Gaps 0;  
QY 1481 ACATGACTGCTGCCCGACAGCTGCTCCAGCGCTGCCCGCCAGCGCTGCCCGCAAGCGCT 1540  
DB 584 ACAACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 643  
QY 1541 GCCCACTGCGGACCGCTCCACCGTGGTCCCACTGCAAGCGCCAGCGCAAGCGCGCGCC 1600  
DB 644 ACTACTCAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 703  
QY 1601 CCATGCCACCTTACAGCGCAATTCCTACTCGGCGCCCGCCAGCGCTGCCCGCCAGCGCT 1660  
DB 704 ACAACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 763  
QY 1661 CTAATACTGCTGCCCGTGAATGCCCAACGACTGTACAGTACAGTACTGCTGCC 1720  
DB 764 ACT 823  
QY 1721 ACCGCAAGCCCTACTACTGCTGCTGCTGAATGCCCAATCATCTGTCATCTAGCGGCTTACC 1780  
DB 824 ACCGCAAGCAACAAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAAC 883  
QY 1781 GTCGCCCGACCGCTGCCCGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1840  
DB 884 ACAACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 943  
QY 1841 GAAATCCCACTACTGTACATCACTACAGTACTGCTGCCCGCACTACCGAGCAAGCTGCC 1900  
DB 944 ACAACACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1003  
QY 1901 CCGACAGCAGTACTGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1960  
DB 1004 ACTGACAGAAACCAAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1063  
QY 1961 AACACAGCAGTACTGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2020  
DB 1064 ACCGCAAGCAACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1123  
QY 2021 ACCAGAGTACTGTACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCT 2080

DB 1124 ACACTACACAGAAACCAACCAACCAACCAACCACTACTACTACTACTACTACTACTACTACT 1183  
QY 2081 GCACCCACACACAGCTGCCCGGTACTACTACTACTACTACTACTACTACTACTACTACTACT 2117  
DB 1184 CCAACCAACCACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1220  
RESULT 8  
ID Q98470 standard; cDNA; 2744 BP.  
AC Q98470;  
DT 01-MAY-1996 (first entry)  
DE Misp1-containing plasmid pMIS1.  
KW Spider silk; repeat unit; consensus: minor ampullate silk protein;  
OS spideroin; Misp; orb web spider; dragline; ds.  
FH Nephila clavipes.  
FH Key Location/Qualifiers  
FT cds 183. .2678  
FT /\*tag= a  
FT /\*tag= b  
FT misc\_difference 996. .1037  
FT /\*tag= b  
FT /note= "represented as indeterminate, a highly  
FT compressed GC rich region which could  
FT not be sequenced"  
PW09252165-A1.  
PD 21-SEP-1995.  
PF 14-MAR-1994; US-209747.  
PR 14-MAR-1994; US-209747.  
PA (UWY-) UNIV WYOMING.  
PI Colgin M, Lewis RV;  
DR WPI; 95-336970/43.  
DR P-PSDB; R80168.  
PT Polypeptide(s) comprising repeated unit amino acid sequences, also  
PT cDNAs - derived from minor ampullate spider silk proteins and used  
PT to form spider silk fibres.  
PS Claim 12; Fig 1; 86pp; English.  
CC cDNA clone, pMIS1, encoding the orb web spider minor ampullate silk  
CC protein Misp1, has been identified and sequenced. Repeat unit peptides  
CC of Misp1 may form part of a larger polypeptide with an amino terminus  
CC (see R80184-85) and a carboxy terminus (see R80186-89). Repeat unit  
CC peptides make up spider silk proteins (spidroins) which in turn  
CC aggregate to form the silk fibres. Spider silk fibres have high  
CC tensile strength and significant elasticity. An isolated cDNA clone of  
CC a silk protein encoding sequence is of use to produce the protein at  
CC high yields using recombinant DNA technology.  
SQ Sequence 2744 BP; 486 A; 549 C; 1059 G; 608 T;  
Query Match 4.3%; Score 120.2; DB 1; Length 2744;  
Best Local Similarity 50.1%; Pred. No. 2,1e-13;  
Matches 333; Conservative 0; Mismatches 323; Indels 9; Gaps 1;  
QY 1482 CAGTACTGCTGCCCGACAGCTGCTCCAGCGCTGCCCGCCAGCGCTGCCCGCAAGCGCTG 1541  
DB 2121 CAGGCGCTCAGCTCTGTCAGCGGCTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2062  
QY 1542 CCCCAGTGGCGACCTCCACCGTGTGCCACTCTGCAAGCGCCAGCGCAATGCAAGCCC 1601  
DB 2061 CTCGGATCTGTCAGCAGCAGCTGTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 2002  
QY 1602 CAGTCCCACTTACAGCGCAATTCCTACTCGGCGCCCGCCAGCGCTGCCCGCCAGCGCAGCTC 1661  
DB 2001 CTTGACCAACCGTACTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCT 1942  
QY 1662 CTACTACTGCTGCCCGTGAATGCCCAACGACTGTACAGTACAGTACTGCTGCTGCCCA 1721  
DB 1941 CAGGCGCTCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1882  
QY 1722 CCGAGCGCCCTACTACTGCTGCTGCTGAATGCCCAATCATCTGTCATCTAGCGGCTTACC 1781  
DB 1881 CAGTCCCGAGCAGCAGCAGCTTGTGCTCGGACCGTATCCAGCAGCAGCAGCAGCAGCAGCTC 1822

QY	1782	CTGCCCCACCGCTGCCCCACCGCTGCCCCACCGAGCCCTTACTACTGTGTGCCAG	1841
Db	1821	CAGTCTCTGCTCTGCACACTGCGACGGGCTTCAGCTCCAGACACTGTGACAGCTCCAGCTC	1762
QY	1842	AAATCCCACTACTGTCACATCACCACACTACTGTGCCCCCACTACCGGACGACCTTGCC	1901
Db	1761	CAGCACGACGACCTCTACGCTAAACGACACACTCTGCTCTGTCACACGAGGACGACGCG	1702
QY	1902	CCAAACACCACTACTGTACTGTATCCACACCACTGTGTGCCCTTACTACCGCAGCA-----C	1952
Db	1701	CACGAGGTCTCTGCTCCGGACACCATAAACGGCTTGTCCACATATACAGACACCAACCGGTGC	1642
QY	1953	CTGTCTCCCAACACCACTAGTACGACCAACCAACGCGACCCCTTACTACCGGACGACCTG	2012
Db	1641	CTGCACTATGCCCCAGCAGCGGGCACACAGCTCCAGCACCAACACTTACCGTAAACCAACG	1582
QY	2013	CCCCCAACACACAGTCACTGTATCCACCACTGTGCCCCCACTGCAAGTCTCCCTTACCG	2072
Db	1581	CGCTCTCGGGTCTGTGTGAGCAGCTGACAGCAACCACTCTCTGTCGGGAGCCGATTC	1522
QY	2073	TGCGCACTGCAACCAACACCACTGTGCCCGGTAACTCAAAACGACGACCACTATCA	2132
Db	1521	CACCTTGAACCAACCGTACCTCCAGAGCTCTGCGGCTGTCAGACAGCTCCAGCTCCAGCAACG	1462
QY	2133	CACCT 2137	
Db	1461	CACCT 1457	

RESULT	9
ID	051557/c
AC	051557; standard: DNA; 6530 BP.
DT	17-MAY-1994 (first entry)
DE	Loricrin gene.
KW	Loricrin; gene; constitutive; inducible; vector; 5' flanking region;
KW	intron; 3' flanking region; TATA box; cap site; intron/exon boundary;
KW	DNA cassette; linker; restriction site; induction; human; psoriasis;
KW	epidermal cell; wound healing; ulcers; growth factor; antisense RNA;
KW	transforming growth factor beta; cytokine; skin; tracheal epithelia;
KW	cancer; E6; E7; papilloma virus; p53 protein; vaccine; oral epithelia;
KW	vaginal epithelia; squamous epithelium; viral; oesophageal epithelia;
KW	corneal epithelia; transduction; transplant; ss.
OS	Homo sapiens.
FH	Key
FT	misc_rna
FT	location/Qualifiers
FT	1..1540
FT	/+tag= a
FT	/note= "5',flanking region"
FT	1387..2677
FT	/+tag= b
FT	/note= "intron and intron/exon boundary"
FT	4384..6530
FT	/+tag= c
FT	/note= "3', flanking region"
PN	M09322431-A..
PD	11-NOV-1993.
PF	28-APR-1993; 003993.
PR	30-APR-1992; US-876286.
PA	(BAYU ) BAYLOR COLLEGE MEDICINE.
PI	Greenhalgh DA, Koop DR, Rothnagel JA;
DR	WPI: 93-368790/46.
PT	Constitutive and inducible vectors based on loricrin or keratin
PT	K6 genes - for selective expression in the epidermis, e.g. for
PT	wound healing, treating psoriasis and skin cancer, etc.
PS	Claim 23; Page 26-32; 68pp; English.
CC	This sequence represents the loricrin gene and was used in the
CC	construction of the loricrin constitutive vector of the invention.
CC	The loricrin vector has a 5' flanking region of approx. 1.5 kb, an
CC	intron of approx. 1.1 kb and a 3' flanking region of approx 1.5 kb.
CC	The 5' flanking region of the loricrin gene comprises a TATA box,
CC	cap site, first intron and intron/exon boundary, all in the correct
CC	order and position for expresion as a nucleic acid cassette. The
CC	loricrin vector also comprises a 3' flanking region of the loricrin

CC gene and a linker with a unique restriction site at the position of  
CC the start and stop codons, joining the two flanking sequences and  
CC also having a position for insertion of the DNA cassette (see also  
CC Q51536). Vectors such as this can be used for: (a) in vitro induction  
CC of human epidermal cells, especially for stimulating healing of  
CC wounds, surgical incisions and ulcers, (where the vector expresses a  
CC growth factor), (b) treating psoriasis, (the vector expressing  
CC transforming growth factor beta, a soluble cytokine or antisense  
CC RNA), (c) treating skin cancer, (the vector expressing antisense RNA  
CC of the E6 or E7 gene of human papilloma virus or normal p53 protein),  
CC (d) for vaccination (the vector expresses a viral capsid protein,  
CC especially of human papilloma virus), or (e) treating cancer of the  
CC squamous epithelium (of epidermal, oral, oesophageal, vaginal,  
CC tracheal or corneal epithelia, the vector expresses antisense RNA).  
CC Alternatively human epithelial cells are transduced *ex vivo*, then  
CC transplanted.  
CC Sequence 6530 BP, 1635 A; 1507 C; 1696 G; 1692 T;

Query Match	3.9%	Score 110.544:	DB 1:	Length 6530:
Best Local Similarity	48.4%:	Pred. No. 1.4e-11:		
Matches 307:	Conservative 0:	Mismatches 327:	Indels 0:	Gaps 0:
QY	1500	CAGCTGCTCCAAACGGCTGCCCCCAACGGGTGCCTCCCAACGGCTGCCCCAATGCGCACCTT	1559	
Dd	3490	CCGCCGCCGCACCGGAGGAACACTTCGGCAGCTCTGTCGCCGCCGCCACCGGAATAAGCCG	3431	
QY	1560	CCACCCTGGTCCCACCTGCAACGCCACCCGGCAATGCAAGCCCCAGTTCCCACTACAACCG	1619	
Dd	3430	CCGCCGCCGCCGACACTGGATTCCACCGCACCTTGTAAGTACCGCCGCCACAGCTAGAAGCCG	3371	
QY	1620	CAATTCTTACTTCGGGCCCCCACACCGGTGGCCCCCACCGCAGTCTTACTACTGTGCCCCCTG	1679	
Dd	3370	CCGGCTGCTTCCACCGCCGCCGGGAATAGCCGCTTCGGCAGCTGGAGCCACACCGGCCACCA	3311	
QY	1680	AATCCCCAACCACTGTACAGTACCACTACTGCTGTCTGCCACCGCAGCCGCCCTTACTATG	1739	
Dd	3310	GAGTACTTGGCCGCTCCGGAAACCGCCGCCGACCAACCGGAGCTTCCGCTTCCAGAGGAG	3251	
QY	1740	CTGTCCCTGAAATCCCAATCATCTGTACATCAGCGCCTAACCGCTGCCGCCACCGCTGCC	1799	
Dd	3250	CCCTCCAGTAGAGACCGCCGCCCTCTGATCCGGCTTCGGCAGCTGGAGCGCGCGCGCG	3191	
QY	1800	CCACCGCTGCCCCCACCGCAGCCCTTACTACTGTGTGCCAGAAATCCCACTACTGTCA	1859	
Dd	3190	CCCCCGGAGGAGCCGCCGCCGAGCTAGAAGCCGCCGCCACCACTTCGGAGTACTTG	3131	
QY	1860	CATACACACACTACGTCTGTCCCTTACCGCAGCACTGGCCCCCAACACCAAGTCACTG	1919	
Dd	3130	ACGCCCCCAACAGACCGCCGACCGCTTCCACAGCTGAGAACCCCGGAGTAGCTGCCG	3071	
QY	1920	TACCAACCACTGCTGCTCCCTACTACTCCGAGACACTGTCTCCCAACACCAAGTACTGAC	1979	
Dd	3070	CCGCCGCCGAGCTGAGAACCCGCCGCTCTCCGGAGTAGCAACCTTCGAGTGGAGCCGCCG	3011	
QY	1980	CACCACCGCAGACCCCTTACTACCGCAGCAGCACTGCCCCCAACCAACAGTCACTGTACAC	2039	
Dd	3010	CCGCCCCCAAGATAGCAACCGCCGAGTAGAGCCACACCGCTTCGGAGTAGCGCGCG	2951	
QY	2040	CCACTGCTGCCCCCACTGCACTCCCCCTTACCCTGTGCACTGCACTGCAACCAACCAAGCTG	2099	
Dd	2950	CCGAGCTAGAGACCAACCCGCCGCTTCGGAGTAGTACTGAGGCCCCCAACCGAGCCGCTCCA	2891	
QY	2100	CCCCGTTAACTACAACAGCAGCGACAGGTCACAC	2133	
Dd	2890	CCGCCGAGCTGGAACCACTTCATAGGAACAC	2857	
RESULT	10			
ID	T80043/c			
AC	T80043 standard; DNA: 40875 bp.			
DT	T80043;			
	03-NOV-1997 (first entry)			



DE	Insert from cosmid 109.
KM	Pectinase; yeast; polygalacturonase; phytopathogen; vegetable spoilage;
KW	Fruit spoilage; food processing; fruit juice; cosmid; ss.
OS	Saccharomyces cerevisiae.
FH	Key
FT	Location/Qualifiers
FT	312. .1943
FT	cds
FT	/*tag= a
FT	/note= "indicated in the specification as an open reading
FT	frame, encoding a protein with homology to
FT	S. cerevisiae allantoinase (DAL5)"
FT	3622. .7107
FT	/*tag= b
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	cds
FT	3965. .5188
FT	/*tag= c
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	3461. .4546
FT	/*tag= d
FT	/note= "indicated in the specification as an open reading
FT	frame, encoding a protein with homologies to
FT	pectinases"
FT	5151. .6194
FT	/*tag= e
FT	/note= "indicated in the specification as an open reading
FT	frame, with homology to S. pombe thiamine
FT	repressed gene"
FT	5486. .5845
FT	/*tag= f
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	cds
FT	6386. .7414
FT	/*tag= g
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	8051. .8917
FT	/*tag= h
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	cds
FT	9656. .10552
FT	/*tag= j
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	9996. .10364
FT	/*tag= k
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	10127. .10936
FT	/*tag= l
FT	/note= "indicated in the specification as an open reading
FT	frame, encoding a protein with homology to the
FT	S. cerevisiae mitochondrial genome maintenance
FT	protein (MGM101)"
FT	11161. .11523
FT	/*tag= m
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	13441. .14517
FT	/*tag= n
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	14058. .14360
FT	/*tag= o
FT	/note= "indicated in the specification as an open reading
FT	frame"
FT	14989. .16119
FT	/*tag= p
FT	/note= "indicated in the specification as an open reading

Query Match	3.9%	Score 110.4	DB 1	Length 40875
Best Local Similarity	48.3%	Pred. No. 2.3e-11		
Matches 309	Conservative 0	Mismatches 331	Indels 0	Gaps 0
DB 1491	CTGCCCGACAGACTGCTCCACGCGGTGCCCGCCCAACGGGCTGCGCCCAACGGGCTGCCCGCAACTG	1550		
DB 24579	CTACGCTCCACACGACCTCTACACTTCTACACTTCTACACGACACTGACTCTCTACGACACT	24520		
DB 1551	CGGACCGCTCCACGCGGTGCCCGCACTGACGACCGCCGCAACTGACGACCGCCGCACTGAC	1610		
DB 24519	CTACGCTCCACACGACCTCTACACTTCTACACTTCTACACGACACTGACTCTCTACGACACT	24460		





FT	cd8	1. .4215
FT		/*tag= a
FT	exon	1. .76
FT		/*tag= b
FT		/number= I
FT	exon	77. .199
FT		/*tag= c
FT		/number= II
FT	exon	200. .319
FT		/*tag= d
FT		/number= III
FT	exon	320. .469
FT		/*tag= e
FT		/number= IV
FT	exon	470. .598
FT		/*tag= f
FT		/number= V
FT	exon	599. .3421
FT		/*tag= g
FT		/number= VI
FT	exon	3422. .3499
FT		/*tag= h
FT		/number= VII
FT	exon	3500. .3636
FT		/*tag= i
FT		/number= VIII
FT	exon	3637. .3796
FT		/*tag= j
FT		/number= IX
FT	exon	3797. .3991
FT		/*tag= k
FT		/number= X
FT	exon	3992. .4117
FT		/*tag= l
FT		/number= XI
FT	exon	4118. .4212
FT		/*tag= m
FT		/number= XII
PN	MO9213075-A.	
PD	06-AUG-1992.	
PE	17-JAN-1992; U00433.	
PR	18-JAN-1991; US-643502.	
PR	10-SEP-1991; US-757022.	
PA	(GEMV ) GENETICS INST INC.	
PI	Clark SC, Gesner TC, Hewick RM, Jacobs K, Turner K,	
DR	WPI: 92-284660/34.	
DR	P-PSDB: R26049.	
PT	New human mega-karyocyte stimulating factors - for treating	
PT	immune deficiencies, cancer, exposure to radiation or drugs,	
PT	bacterial and viral infections, etc.	
PS	Claim 1, 2 and 3; Fig 1; 87pp; English.	
CC	The sequence covered by exons II, III and IV encodes megakaryocyte	
CC	stimulating factor (MSF). This sequence is modified by the addition	
CC	of an N-terminal sequence encoding a secretory leader, an initiating	
CC	methionine preceding exon II and a terminating codon following exon	
CC	IV. The cDNA sequence given contains sequences derived from human	
CC	megakaryocyte colony stimulating factor (meg-CSF). Exon I contains	
CC	the initiating methionine, and encodes a classical mammalian protein	
CC	secretion signal sequence. The sequence encoding the original meg-CSF	
CC	includes exons II-IV and is thought to terminate in the region between	
CC	amino acid residues 134 - 147. The primary transcript of this gene	
CC	may be cleaved in different ways to yield a family of mRNA's each	
CC	encoding a different MSF protein. Exons V and VI are thought to be	
CC	related to the activity of the factor and are also implicated in the	
CC	stability, folding and processing of the molecule. These exons are	
CC	also thought to play a role in the observed synergy of MSF with other	
CC	cytokines. Exons V - XII are believed to be implicated in the	
CC	processing or folding of the appropriate structure of the resulting	
CC	factor, ie. one or more of these exons may contain sequences which	
CC	direct proteolytic cleavage, adhesion, organisation of the cellular	
CC	matrix or extracellular matrix processing. Both naturally occurring	
CC	and non-naturally occurring MSF's may be characterised by various	
CC	combinations of alternatively spliced exons from this sequence, with	

Query Match	Best Local Similarity	Score	DB 1;	Length	5008;
Matches 318;	Conservative 0;	Mismatches 362;	Indels 0;	Gaps 0;	
CC the exons spliced together in differing orders to form different CC members of the MSF family.	5008 BP;	1701 A;	1506 C;	797 G;	1004 T;
Sequence					
Query Match	3.6%;	Score 100.8;	DB 1;	Length 5008;	
Best Local Similarity	46.8%;	Pred. No. 7.3e-10;			
Matches 318;	Conservative 0;	Mismatches 362;	Indels 0;	Gaps 0;	
1496 CCGACAGTGTCTCCACACGCGTGGCCCAAGCGGTGGCCCAACGCGTGGCCCAACTGGCCGA	1555				
1096 CCCACACCTTACCATCATTAAGTGTGCACCCACACCCCCCAAGAGCTGTGCACCCACGAC	1155				
1556 CCCTCCACCGTGGTCCCACTGTCAACGGCACCACCGACATGACAGCCCAAGTCCACTTACA	1615				
1156 ACCAGTGTGCACCCACCACTGTCCCAAGAGGCTGTGACCCACCAACCAAGAGAGCTTGA	1215				
1616 ACCGCAATTCTACTTCCGGCCCAACCGGTGGCCCAACCGGACGCTCTACTACTGTGCTGC	1675				
1216 CCCACACCTTCCCAAGAGGCTGTGCACCCACACCAAGAGAGGCTGTGCACCCACGAC	1275				
1676 CCTGAATGCCCAACGACGTGTCACTTACCACTTGTGTGTGTCCCAACGCGCCCTTACT	1735				
1276 AAGTGTGACCCACGACCTTCCCAAGAGGCTGTGACCCACACCCCAAGAGGCTTCCCA	1335				
1736 ACTGTGTCTCCCTGAATTCCTCAATCTGTCATGACGAGCTGACGCTGCTGCCACCGT	1795				
1336 ACTTACCCCAAGAGGCTGTGCACCCACGACCTTCCCAAGAGGCTTACACACCACTTCCAA	1395				
1796 GCCCCACCGCTGTGCCCCACCGACCGCTTACTGTGTGTGTCCCAAAATCCCACTACT	1855				
1396 GAGCTGTGACCCACGACCAAGAGGCTGTGACCCACGACCTTCCCAAGAGGCTTCCCA	1455				
1856 GTTCATATCACCACTTACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT	1915				
1456 GCCCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCA	1515				
1916 ACTGTACACCGACGCTGTGCCCCCTTACTGACGAGACGCTGCTCCCAACACGACGTACT	1975				
1516 CCGTGCACCCACGACCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGG	1575				
1976 GCACGACCCACGACCGCTTACTGACGAGACGCTGTGCCCCCAACGACCACTCTGTA	2035				
1576 ACCACCAAGTGTGACCCACGACCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGG	1635				
2036 CCACGACGCTGTGCCCCCTTACTGACGAGACGCTGTGCCCCCAACGACCACTCTGTA	2095				
1636 CCACGACGCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGG	1695				
2096 GCTGCGCGGTACTATCAACGACGAGCAGTACCACTCTGAGAGGATGATACATCGAC	2155				
1696 AAGGAGGCTTGCACCCACGACCGCTTCCCAAGAGGCTTCCCAAGAGGCTTCCCAAGAGG	1755				
2156 CCCCTTCTCCCAAGGACCC 2175					
1756 CCCACGACCTTCCCAAGGACCC 1775					
RESULT 15					
ID T07199					
AC T07199: standard; DNA; 913 BP.					
DE 10-MAY-1996 (first entry)					
KW H6 gene; promoter; cotton; transgenic plant; fiber; immobilisation;					
OS Gossypium hirsutum strain Coker 312.					
FT Key Location/Qualifiers					
FT cds 72..716					
FT /tag= a					
FT signal_peptide 72..147					
FT /tag= b					
FT mat_peptide 148..713					

FT US5474925-A. /tag= c  
 PN 12-DEC-1995.  
 PD 19-DEC-1991; 812233.  
 PR 19-DEC-1991; US-812233.  
 PR 23-MAR-1994; US-217327.  
 PA (CETU) AGRACETUS INC.  
 PI Barton KA, Maliyakal J;  
 DR WPI; 96-039513/04.  
 P-PSDB; R86913.  
 PT Prepn. of immobilised enzymes in cotton fibre - esp.  
 PT pesticide-degrading enzymes e.g. pectinase hydrolase  
 PS Disclosure; Column 31-34; 37pp; English.  
 CC The cotton H6 gene (T07198) encodes an extensin-like protein  
 CC (R86913) that is an integral part of the secondary cell wall  
 CC of cotton. The portions of the gene encoding the N-terminal  
 CC or transmembrane regions of the protein may be utilised in the  
 CC construction of gene fusions used to anchor (i.e. immobilise)  
 CC foreign proteins in the fibers of transgenic cotton plants.  
 CC Fusions of H6 and a foreign protein should be localised within  
 CC the secondary walls of the fiber. The immobilised proteins,  
 CC e.g. enzymes or antibodies, have industrial, medical and  
 CC environmental applns.  
 SQ Sequence 913 bp; 196 A; 313 C; 155 G; 249 T;

Query Match 3.5%; Score 100; DB 1; Length 913;  
 Best Local Similarity 51.8%; Pred. No. 6.9e-10;  
 Matches 226; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

OY 1581 CGCCACCCGACAGTGGAGCCCGGACCTTCAACCGCAATTCCTACTCCGGCCCCA 1640  
 DB 175 CGACACCCGACACCGGCTTCTACTCTCTCCGACGACTCAAGCACCGCTTACACAA 234  
 OY 1641 CGGCTGCCCCGACCGGCTTCTACTACTGCTGCGCCCTGAATCCCAACCACTGTACAG 1700  
 DB 235 CGGCTGCCCCGACCGGCTTCTACTCTCTCCGACGACTTCAATCAGCGCCCGAGTACAG 294  
 OY 1701 TACACCTACTGCTGCTCCACCGGACCGCTTCTACTGCTGCTGCTGAATCCCAATCA 1760  
 DB 295 CTTCCTCAGCCCGGAGTTTCAATCTCCACCGGCTTCTCTCTGCAATCTCCACGAC 354  
 OY 1761 CTGTCAATCAGGCGCTTACCGCTGCGCCCGGACCGCTGCGCCCGGACCGGACG 1820  
 DB 355 CTGCTTCTCTCTCTCTGCAATCTCCACGCTTCTCCACCTCTCTGCACTCTCTCCAC 414  
 OY 1821 CCCCTACTGCTGCTGCTCCGAAATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTG 1880  
 DB 415 CAGCTTCTCAGCTTCTCCGACCTTCTCCACGCTTCTCCACGCTTCTCCACGCTTCTC 474  
 OY 1881 CGACTACCGGACGCTGCGCCCGGACGACGACTGCTGCTGCTGCTGCTGCTGCTGCT 1940  
 DB 475 CTGCTACCGGACGCTGCGCCCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 534  
 OY 1941 CTACCGGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2000  
 DB 535 CAGTACAAACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 594  
 OY 2001 CCGACGACCTGCGCC 2016  
 DB 595 TCGGGGCTGCTAGGCC 610

Search completed: July 18, 1999, 11:08:38  
 Job time: 2188 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 11:07:32 ; Search time 1043.53 Seconds

(without alignments)  
3588.701 Million cell updates/sec

Title: US-09-294-663-1

Perfect score: 2455

Sequence: 1 GTAACGTTAAGTGAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY\_NUC

Searched: 2002476 segs, 762712212 residues

Database :

EST:\*  
1: em\_est1:\*  
2: em\_est2:\*  
3: em\_est3:\*  
4: em\_est4:\*  
5: em\_est5:\*  
6: em\_est6:\*  
7: em\_est7:\*  
8: em\_est8:\*  
9: em\_est9:\*  
10: em\_est10:\*  
11: em\_est11:\*  
12: em\_est12:\*  
13: em\_est13:\*  
14: em\_est14:\*  
15: em\_est15:\*  
16: em\_est16:\*  
17: em\_est17:\*  
18: em\_est18:\*  
19: em\_est19:\*  
20: em\_est20:\*  
21: em\_est21:\*  
22: em\_est22:\*  
23: em\_est23:\*  
24: em\_est24:\*  
25: em\_est25:\*  
26: em\_est26:\*  
27: em\_est27:\*  
28: em\_est28:\*  
29: em\_est29:\*  
30: em\_est30:\*  
31: em\_est31:\*  
32: em\_est32:\*  
33: em\_est33:\*  
34: em\_est34:\*  
35: em\_est35:\*  
36: em\_est36:\*  
37: em\_est37:\*  
38: em\_est38:\*  
39: em\_est39:\*  
40: em\_est40:\*  
41: em\_est41:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

# SUMMARIES

Result	No.	Score	Query Match	Length	ID	Description
c	1	110	4.5	538	30	A114969 u141b06.y
	2	107.8	4.4	814	30	A1131654 u141b06.x

RESULT	LOCUS	DEFINITION	ACCESSION	NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
1	A114969/c	u141b06.y1 Sugano mouse embryo mewa Mus musculus cDNA clone 1884947 5' similar to TR:014633 O14633 SKIN-SPECIFIC PROTEIN. ; mRNA sequence.	A114969	93515293	EST.	House mouse. Mus musculus	Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	1 (bases 1 to 538)	Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisler, S., Kucada, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.
									The WashU-HMI Mouse EST Project Unpublished (1996)

## ALIGNMENTS

TITLE	JOURNAL	COMMENT
Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine		









NID	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
92402857	EST.	house mouse.	Eukaryotae; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;		
Mus musculus			Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
1 (bases 1 to 452)					
Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucada,T., Lacy,M., Le,M., Martin,J., Morris,B., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Weston,R.					
The WashU-HMI Mouse EST Project					
Unpublished (1996)					
TITLE	JOURNAL	COMMENT			
Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108 Tel.: 314 286 1800 Fax: 314 286 1810 Email: mouseest@wustl.wustl.edu This clone is available royalty-free through LNL ; contact the IMAGE Consortium (info@image.llnl.gov) for further information. MGI:556135 Possible reversed clone: polyT not found Seq primer: -28m13 rev1 ET from Amersham High quality sequence stop: 447. Location/Qualifiers 1..452 /organism="Mus musculus" /strain="C57BL/6" /note="Organ: skin; Vector: pBluescript SK-; Site.1: EcoRI; Site.2: XhoI; Cloned unidirectionally. Primer: Oligo dT. Whole skin from 11 week old C57BL/6 female mice Average insert size: 1.0 kb; Uni-ZAP XR Vector; -5' adaptor sequence: 5' CTCGAGGTCTTTTCTTTTTTTT 3' /db_xref="taxon:10090" /clone="975431" /clone.lib="Stratagene mouse skin (#937313)" /sex="females" /tissue_type="whole skin" /dev_stage="11 weeks old" /lab_host="SOLR (kanamycin resistant)" complement( <L...>452 )					
BASE COUNT	32 a	93 c	193 g	134 t	
ORIGIN					
Query Match	3.7%	Score 90.2;	DB 22;	Length 452;	
Best Local Similarity	54.3%	Pred. No. 6.9e-11;			
Matches 182;	Conservative	0;	Mismatches 155;	Indels 0;	Gaps 0;
OY	1515	CTGCTCCAAAGCGCTGCCCAACGGCTGCACCGCTGCCCAACTGCGGACCGCTCCA	1574		
Db	346	CTTTGCTGCAACACCCTTCCTCCACAGCCAGCCGACACAGGACGACACAGCACAG	287		
OY	1575	CCGCTGCTCCACTTCGAACGCCACCCGCGACAATGCACTCCCACTACAAACCGCAA	1634		
Db	286	GAGCGCGACACACACGGGGGACGAGCACTACCACTGGCTGAGGACGACACAGCA	227		
OY	1635	TTCTCTACTCTCGGCCCCACACCGCTGCCGCCACGCGAGCTCTACTACTGTGCGCCGTA	1694		
Db	226	CCACAGCAGCCACCAACAAAGCCACAGCCACGACCAACCAACAAAGCAGCCGCGA	167		
OY	1655	CCCACAACCACTGTACAGACTACTGTCTGCTCCACGCGAGCCCTACTACTGCTG	1754		
Db	166	CCACAGCCACCCAGCCACAGCCACACGACCAAGCCACACAGCCACAGCCAGCGAG	107		
OY	1755	TCCCGAATATCCCATGTCATGTCATGAGGCGTATGCGTGCCGCCACGCGTGC	1814		
Db	106	CCACAGCCACCAACGACCAACACGACGCGACGCTGCTGACAGCGCGCCACAGCA	47		

Query Match	Best Local Similarity	Matches	Score	DB	Length	514
Conservative	54.3%	0	Pred. No. 7e-11	Mismatches	153	Indels
0	0	0	0	0	0	0
1515	CTGCTCCAAAGGCTGCCCAACGAGGCTGCGCCCACTGCGGACCTTCA	1574				
Db	332	CTTTGTGCAACACCTTCCACAGCAGCGCAGCCACAGAGCCACAGCAG	293			
Qy	1575	CGGTGCGCCACTGCAAGCCAGCCAGCAATGACGCCAGTCCACCTACACCGCA	1634			
Db	292	GAGCGGACAGCAGCAGCGGCGGACGACGACACTACCATGCTTGCAGCAGCAGCAG	233			
Qy	1635	TTTCTACTTCTGCGGCCACCGCTGCCCCACCGCAGCTCTACTAGCTGCGCCCTGAAT	1694			











GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 01:18:06 : Search time 1514.49 Seconds  
(without alignments)  
3603.202 Million cell updates/sec

Title: US-09-294-663-1

Perfect score: 2455  
Sequence: 1 GTACGCTAGTGAAGAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY\_NNC

Searched: 3232953 seqs, 1111407957 residues

Database :

Pending\_Patents\_NA:\*  
1: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*  
2: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*  
3: /cgn2\_6/ptodata/1/pna/US080\_COMB.seq:\*  
4: /cgn2\_6/ptodata/1/pna/US081A\_COMB.seq:\*  
5: /cgn2\_6/ptodata/1/pna/US081B\_COMB.seq:\*  
6: /cgn2\_6/ptodata/1/pna/US081C\_COMB.seq:\*  
7: /cgn2\_6/ptodata/1/pna/US081D\_COMB.seq:\*  
8: /cgn2\_6/ptodata/1/pna/US082A\_COMB.seq:\*  
9: /cgn2\_6/ptodata/1/pna/US082B\_COMB.seq:\*  
10: /cgn2\_6/ptodata/1/pna/US082C\_COMB.seq:\*  
11: /cgn2\_6/ptodata/1/pna/US083A\_COMB.seq:\*  
12: /cgn2\_6/ptodata/1/pna/US083B\_COMB.seq:\*  
13: /cgn2\_6/ptodata/1/pna/US084A\_COMB.seq:\*  
14: /cgn2\_6/ptodata/1/pna/US084B\_COMB.seq:\*  
15: /cgn2\_6/ptodata/1/pna/US084C\_COMB.seq:\*  
16: /cgn2\_6/ptodata/1/pna/US084D\_COMB.seq:\*  
17: /cgn2\_6/ptodata/1/pna/US085\_COMB.seq:\*  
18: /cgn2\_6/ptodata/1/pna/US086\_COMB.seq:\*  
19: /cgn2\_6/ptodata/1/pna/US087A\_COMB.seq:\*  
20: /cgn2\_6/ptodata/1/pna/US087B\_COMB.seq:\*  
21: /cgn2\_6/ptodata/1/pna/US087C\_COMB.seq:\*  
22: /cgn2\_6/ptodata/1/pna/US088A\_COMB.seq:\*  
23: /cgn2\_6/ptodata/1/pna/US088B\_COMB.seq:\*  
24: /cgn2\_6/ptodata/1/pna/US088C\_COMB.seq:\*  
25: /cgn2\_6/ptodata/1/pna/US089A\_COMB.seq:\*  
26: /cgn2\_6/ptodata/1/pna/US089B\_COMB.seq:\*  
27: /cgn2\_6/ptodata/1/pna/US089C\_COMB.seq:\*  
28: /cgn2\_6/ptodata/1/pna/US089D\_COMB.seq:\*  
29: /cgn2\_6/ptodata/1/pna/US090A\_COMB.seq:\*  
30: /cgn2\_6/ptodata/1/pna/US090B\_COMB.seq:\*  
31: /cgn2\_6/ptodata/1/pna/US090C\_COMB.seq:\*  
32: /cgn2\_6/ptodata/1/pna/US090D\_COMB.seq:\*  
33: /cgn2\_6/ptodata/1/pna/US091A\_COMB.seq:\*  
34: /cgn2\_6/ptodata/1/pna/US091B\_COMB.seq:\*  
35: /cgn2\_6/ptodata/1/pna/US091C\_COMB.seq:\*  
36: /cgn2\_6/ptodata/1/pna/US091D\_COMB.seq:\*  
37: /cgn2\_6/ptodata/1/pna/US092A\_COMB.seq:\*  
38: /cgn2\_6/ptodata/1/pna/US092B\_COMB.seq:\*  
39: /cgn2\_6/ptodata/1/pna/US092C\_COMB.seq:\*  
40: /cgn2\_6/ptodata/1/pna/US092D\_COMB.seq:\*  
41: /cgn2\_6/ptodata/1/pna/US093A\_COMB.seq:\*  
42: /cgn2\_6/ptodata/1/pna/US093B\_COMB.seq:\*  
43: /cgn2\_6/ptodata/1/pna/US093C\_COMB.seq:\*  
44: /cgn2\_6/ptodata/1/pna/US093D\_COMB.seq:\*  
45: /cgn2\_6/ptodata/1/pna/US094A\_COMB.seq:\*  
46: /cgn2\_6/ptodata/1/pna/US094B\_COMB.seq:\*  
47: /cgn2\_6/ptodata/1/pna/US094C\_COMB.seq:\*  
48: /cgn2\_6/ptodata/1/pna/US094D\_COMB.seq:\*  
49: /cgn2\_6/ptodata/1/pna/US095A\_COMB.seq:\*  
50: /cgn2\_6/ptodata/1/pna/US095B\_COMB.seq:\*  
51: /cgn2\_6/ptodata/1/pna/US095C\_COMB.seq:\*  
52: /cgn2\_6/ptodata/1/pna/US095D\_COMB.seq:\*  
53: /cgn2\_6/ptodata/1/pna/US096A\_COMB.seq:\*

54: /cgn2\_6/ptodata/1/pna/US6009A\_COMB.seq:\*  
55: /cgn2\_6/ptodata/1/pna/US6009B\_COMB.seq:\*  
56: /cgn2\_6/ptodata/1/pna/US6010A\_COMB.seq:\*  
57: /cgn2\_6/ptodata/1/pna/US6010B\_COMB.seq:\*  
58: /cgn2\_6/ptodata/1/pna/US6011\_COMB.seq:\*  
59: /cgn2\_6/ptodata/1/pna/US6012\_COMB.seq:\*  
60: /cgn2\_6/ptodata/1/pna/US6013\_COMB.seq:\*  
61: /cgn2\_6/ptodata/1/pna/PCF\_NEW\_COMB.seq:\*  
62: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
63: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
64: /cgn2\_6/ptodata/1/pna/US08\_NEW\_COMB.seq:\*  
65: /cgn2\_6/ptodata/1/pna/US60\_NEW\_COMB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2455	100.0	2455	US-09-103-429-1	Sequence 1, Appl 1
2	2455	100.0	2455	US-09-103-429A-1	Sequence 1, Appl 1
3	2455	100.0	2455	US-09-294-663-1	Sequence 1, Appl 1
4	2354.4	95.9	2821	US-09-103-429-2	Sequence 2, Appl 1
5	2354.4	95.9	2821	US-09-103-429A-2	Sequence 2, Appl 1
6	2354.4	95.9	2821	US-09-294-663-2	Sequence 2, Appl 1
7	153.6	6.3	799	US-08-733-365A-105	Sequence 105, App
8	153.6	6.3	799	US-08-733-365A-104	Sequence 104, App
9	153.4	6.2	803	US-08-970-900-7	Sequence 7, Appl 1
10	153.4	6.2	803	US-08-733-365A-112	Sequence 112, App
11	153.2	6.2	798	US-08-970-900-4	Sequence 4, Appl 1
12	151.6	6.2	1056	US-08-407-128-15	Sequence 15, Appl 1
13	151.6	6.2	1056	US-08-407-128-15	Sequence 15, Appl 1
14	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
15	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
16	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
17	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
18	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
19	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
20	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
21	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
22	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
23	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
24	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
25	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
26	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
27	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
28	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
29	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
30	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
31	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
32	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
33	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
34	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
35	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
36	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
37	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
38	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
39	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
40	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
41	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
42	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
43	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
44	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1
45	151.6	6.2	1056	US-08-487-823-15	Sequence 15, Appl 1

#### ALIGNMENTS

RESULT 1



```

US-09-103-429-1
: Sequence 1, Application US/09103429
: GENERAL INFORMATION:
:   APPLICANT: Granados, Robert R
:   TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin
:   TITLE OF INVENTION: CDNA and Related Products and Methods
:   NUMBER OF SEQUENCES: 2
:   CORRESPONDENCE ADDRESS:
:     ADDRESSEE: Brown, Pinnisi & Michaels, P.C.
:     STREET: 118 North Riosa
:     CITY: Ithaca
:     STATE: NY
:     COUNTRY: USA
:     ZIP: 14850
: COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentin Release #1.0, Version #1.30
:   CURRENT APPLICATION DATA:
:     APPLICATION NUMBER: US/09/103,429
:     FILING DATE: 24-JUN-1998
:     CLASSIFICATION: 800
:   ATTORNEY/AGENT INFORMATION:
:     NAME: Michaels, Christopher A
:     REGISTRATION NUMBER: 34,390
:     REFERENCE/DOCKET NUMBER: BTI-39
:     TELECOMMUNICATION INFORMATION:
:       TELEPHONE: (607) 256-2000
:       TELEFAX: (607) 256-3628
:     INFORMATION FOR SEQ ID NO: 1:
:       SEQUENCE CHARACTERISTICS:
:         LENGTH: 2455 base pairs
:         TYPE: nucleic acid
:         STRANDEDNESS: double
:         TOPOLOGY: linear
:       MOLECULE TYPE: cDNA
:       HYPOTHETICAL: NO
:       ANTI-SENSE: NO
:       FRAGMENT TYPE: N-terminal
:       ORGANISM: Trichoplusia ni
:       TISSUE TYPE: Peritrophic Membrane
:       IMMEDIATE SOURCE:
:       CLONE: IM14
: US-09-103-429-1

```

Query Match 100.0%; Score 2455; DB 33; Length 2455;  
 Best Local Similarity 100.0%; Pred. No. 0;  
 Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 GTACGTTAAGTGAAGAAATTAACAGGACAACTTATGATTAAGACCCCTCTTCTTCT 60
D 1 GTACGTTAAGTGAAGAAATTAACAGGACAACTTATGATTAAGACCCCTCTTCTTCT 60
QY 61 GACGGCCCTCGGGCTCGTGGCGGCGCTCTGAGTCAAGCGGCGGGAAGAAGACCCGC 120
D 61 GACGGCCCTCGGGCTCGTGGCGGCGCTCTGAGTCAAGCGGCGGGAAGAAGACCCGC 120
QY 121 TCTCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
D 121 TCTCAGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 180
QY 181 ATAGACATGACACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCAATGCGACCGAGAGA 240
D 181 ATAGACATGACACCAAGTTCTACTACTGTGAATATGGTCTCAAGTTCAATGCGACCGAGAGA 240
QY 241 CTGTGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
D 241 CTGTGCTCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300

```

```

QY 301 CGATGACACCTGCGAGAGACCTGACGCTGAGACAAACCCAGGCCCGGAGAACACTCAGGC 360
D 301 CGATGACACCTGCGAGAGACCTGACGCTGAGACAAACCCAGGCCCGGAGAACACTCAGGC 360
QY 361 CCCAACAACACCCAGGCGCCCAACACACACTACTCAGGCGCCCTACTACACACCCAGGC 420
D 361 CCCAACAACACCCAGGCGCCCAACACACACTACTCAGGCGCCCTACTACACACCCAGGC 420
QY 421 CCCAACAACACCCAGGCGCCCAACACACACTACTCAGGCGCCCTACTACACACCCAGGC 480
D 421 CCCAACAACACCCAGGCGCCCAACACACACTACTCAGGCGCCCTACTACACACCCAGGC 480
QY 481 AACTACACCTGAGGCGCCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 540
D 481 AACTACACCTGAGGCGCCCTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 540
QY 541 TACCAACAACACCCAGGCGCCCAACACACACTACTACTACTACTACTACTACTACTACTACT 600
D 541 TACCAACAACACCCAGGCGCCCAACACACACTACTACTACTACTACTACTACTACTACTACT 600
QY 601 TACCAACAACACCCAGGCGCCCAACACACACTACTACTACTACTACTACTACTACTACTACT 660
D 601 TACCAACAACACCCAGGCGCCCAACACACACTACTACTACTACTACTACTACTACTACTACT 660
QY 661 GCGCGCAACTACCCCGGCGGCACTACCCCTGCGCGCAACCCCGGCGGCACTACCC 720
D 661 GCGCGCAACTACCCCGGCGGCACTACCCCTGCGCGCAACCCCGGCGGCACTACCC 720
QY 721 AGGTGTTCTGACACCCACTTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
D 721 AGGTGTTCTGACACCCACTTACGCGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 780
QY 781 TGGTGGCCCACTGACCTGACATGCACTGATGATGATGATGATGATGATGATGATGATGATG 840
D 781 TGGTGGCCCACTGACCTGACATGCACTGATGATGATGATGATGATGATGATGATGATGATG 840
QY 841 CTCTACACGTTGCTCAACGGTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 900
D 841 CTCTACACGTTGCTCAACGGTTACCTTACCTTACCTTACCTTACCTTACCTTACCTTACCTT 900
QY 901 CAACCCCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
D 901 CAACCCCTGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 960
QY 961 CCCGCGACCCCACTGACAGAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1020
D 961 CCCGCGACCCCACTGACAGAGGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1020
QY 1021 CGACAATGATGCGCAGCTTAAGTGAATGAGTGGCTTGGCCCGGCAAGCCGTTG 1080
D 1021 CGACAATGATGCGCAGCTTAAGTGAATGAGTGGCTTGGCCCGGCAAGCCGTTG 1080
QY 1081 CGACAATGATGCGCAGCTTAAGTGAATGAGTGGCTTGGCCCGGCAAGCCGTTG 1140
D 1081 CGACAATGATGCGCAGCTTAAGTGAATGAGTGGCTTGGCCCGGCAAGCCGTTG 1140
QY 1141 CCACCTGAGTTTGAATGCAATGATGCAATGATGCAATGATGCAATGATGCAATGATGCAATG 1200
D 1141 CCACCTGAGTTTGAATGCAATGATGCAATGATGCAATGATGCAATGATGCAATGATGCAATG 1200
QY 1201 CGGCGGCGAGAGCGAAGATTGAGCTGAGAGAGATGCGCAACCGGCTGTAAGTGGCC 1260
D 1201 CGGCGGCGAGAGCGAAGATTGAGCTGAGAGAGATGCGCAACCGGCTGTAAGTGGCC 1260
QY 1261 CAGGGAACCAATGATGATGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
D 1261 CAGGGAACCAATGATGATGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1320
QY 1321 CCACCTGAGTTTGAATGCAATGATGCAATGATGCAATGATGCAATGATGCAATGATGCAATG 1380
D 1321 CCACCTGAGTTTGAATGCAATGATGCAATGATGCAATGATGCAATGATGCAATGATGCAATG 1380
QY 1381 TATGCGAAGACCTTGGCTTGAAGACCTGCACTGATGCTGCGCACAGTCTGTGAGTC 1440

```

|||||  
Db 1381 TATGCAAAACCTTGCCCTGGAAACCTGCATCTCCTGCGCACAGCTCGTAGTC 1440  
|||  
Qy 1441 TCCTGTGACCGCTGGTGGCAAGTTTTCGAGTTCGATTCGACACACAGTGCATCGAC 1500  
|||  
Db 1441 TCCTGTGACCGCTGGTGGCAAGTTTTCGAGTTCGATTCGACACACAGTGCATCGAC 1500  
|||  
Qy 1501 TCGTGGCCCGGACAGCTGCTCCAAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAG 1560  
|||  
Db 1501 TCGTGGCCCGGACAGCTGCTCCAAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAG 1560  
|||  
Qy 1561 TCGCGGACCTTCACCGCTGGTCCGACCTGACAGCCGACCGGACGCTGACGCTCCG 1620  
|||  
Db 1561 TCGCGGACCTTCACCGCTGGTCCGACCTGACAGCCGACCGGACGCTGACGCTCCG 1620  
|||  
Qy 1621 ACCTACACCGCAATCTCTACTCGGCGCCGACGCTGGCGCCGACGCTGCTCTACTAC 1680  
|||  
Db 1621 ACCTACACCGCAATCTCTACTCGGCGCCGACGCTGGCGCCGACGCTGCTCTACTAC 1680  
|||  
Qy 1681 TCGTGGCCCGTGAATCCCAACAGCTGTCACAGTACACCTACTGCTGCCACGCGACG 1740  
|||  
Db 1681 TCGTGGCCCGTGAATCCCAACAGCTGTCACAGTACACCTACTGCTGCCACGCGACG 1740  
|||  
Qy 1741 CCTCTACTGCTGCTGCTCCGGAATCCCAATCTCTACTGCTGCTGCTGCTGCTGCTG 1800  
|||  
Db 1741 CCTCTACTGCTGCTGCTCCGGAATCCCAATCTCTACTGCTGCTGCTGCTGCTGCTG 1800  
|||  
Qy 1801 CACCGCTGGCCCGGACGCTGGCGCCGACGCTGGCGCCGACGCTGCTGCTGCTGCTG 1860  
|||  
Db 1801 CACCGCTGGCCCGGACGCTGGCGCCGACGCTGGCGCCGACGCTGCTGCTGCTGCTG 1860  
|||  
Qy 1861 AACTACTGTCACATCACACCTACTGCTGCCCGACGCTGCTGCCCGACGCTGCTGCC 1920  
|||  
Db 1861 AACTACTGTCACATCACACCTACTGCTGCCCGACGCTGCTGCCCGACGCTGCTGCC 1920  
|||  
Qy 1921 CACAGTCACTGTAACACCGCAGCTGCTGCCCGACGCTGCTGCCCGACGCTGCTGCC 1980  
|||  
Db 1921 CACAGTCACTGTAACACCGCAGCTGCTGCCCGACGCTGCTGCCCGACGCTGCTGCC 1980  
|||  
Qy 1981 AGTCACTGTAACACCGCAGCTGCTGCCCGACGCTGCTGCCCGACGCTGCTGCCCGAC 2040  
|||  
Db 1981 AGTCACTGTAACACCGCAGCTGCTGCCCGACGCTGCTGCCCGACGCTGCTGCCCGAC 2040  
|||  
Qy 2041 CACACACACAGCTGCTGCCCGGTAACACACGACGCTGCTGCCCGACGCTGCTGCCCG 2100  
|||  
Db 2041 CACACACACAGCTGCTGCCCGGTAACACACGACGCTGCTGCCCGACGCTGCTGCCCG 2100  
|||  
Qy 2101 TGAATCGACCCCTCTCTCCCAACGACCCCATCAACCTTGGCTTGAAGAAATGCAACGT 2160  
|||  
Db 2101 TGAATCGACCCCTCTCTCCCAACGACCCCATCAACCTTGGCTTGAAGAAATGCAACGT 2160  
|||  
Qy 2161 TTTGGCCCTGGGCTCAGCGTGCATGCGCAAAATCTGGGCTGAGAGGCAACCAAGT 2220  
|||  
Db 2161 TTTGGCCCTGGGCTCAGCGTGCATGCGCAAAATCTGGGCTGAGAGGCAACCAAGT 2220  
|||  
Qy 2221 CCTGTGCTGCTGCTGAGAGGCTCTCCAGTTCAACCCCACTACTAGACCTGTGACTTCG 2280  
|||  
Db 2221 CCTGTGCTGCTGCTGAGAGGCTCTCCAGTTCAACCCCACTACTAGACCTGTGACTTCG 2280  
|||  
Qy 2281 TTTGCAACGTCGCTGAGAGGCTCTCCAGTTCAACCCCACTACTAGACCTGTGACTTCG 2340  
|||  
Db 2281 TTTGCAACGTCGCTGAGAGGCTCTCCAGTTCAACCCCACTACTAGACCTGTGACTTCG 2340  
|||  
Qy 2341 TTTGCAACGTCGCTGAGAGGCTCTCCAGTTCAACCCCACTACTAGACCTGTGACTTCG 2400  
|||  
Db 2341 TTTGCAACGTCGCTGAGAGGCTCTCCAGTTCAACCCCACTACTAGACCTGTGACTTCG 2400  
|||  
Qy 2401 GTAAACCTACTTAATTAATGAAGGTTTGTATTAATTAATTAATTAATTAATTAATTA 2455  
|||  
Db 2401 GTAAACCTACTTAATTAATGAAGGTTTGTATTAATTAATTAATTAATTAATTAATTA 2455  
|||

RESULT 2

US-09-103-429A-1  
; Sequence 1, Application US/09103429A  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
; STREET: 118 North Tioja  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429A  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2455 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; TISSUE TYPE: peritrophic Membrane  
; IMMEDIATE SOURCE:  
; CLONE: IM14  
; US-09-103-429A-1  
  
Query Match 100.0%; Score 2455; DB 33; Length 2455;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
  
Qy 1 GTAACTTAAGTAAAGATAAACAAGCAAGTATGATTAAGACCTCTATTCT 60  
Db 1 GTAACTTAAGTAAAGATAAACAAGCAAGTATGATTAAGACCTCTATTCT 60  
|||  
Qy 61 GAGGCGCTCGGCTGCTGCGCGGCTCTGAAGTCAGCGACGCGAGAAACCCGCG 120  
Db 61 GAGGCGCTCGGCTGCTGCGCGGCTCTGAAGTCAGCGACGCGAGAAACCCGCG 120  
|||  
Qy 121 TCTCCAGAGCGGACCCGACGCTGCGCTGCGCTGAGAGCACTGCGCTGCTGCTGCTG 180  
Db 121 TCTCCAGAGCGGACCCGACGCTGCGCTGCGCTGAGAGCACTGCGCTGCTGCTGCTG 180  
|||  
Qy 181 ATACGACTGACCAAGTCTACTACTGTAATATGCTCAAGTTCAATGACCGAGAGA 240  
Db 181 ATACGACTGACCAAGTCTACTACTGTAATATGCTCAAGTTCAATGACCGAGAGA 240  
|||  
Qy 241 CTGTGCTCTGTATGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
Db 241 CTGTGCTCTGTATGCAAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 300  
|||

QY 301 CGGATGACCCCTGCGAGGACCTCCAGCTGAGACAAACCCAGGCCCGACAACTCAGGC 360  
| | | | |  
Db 301 CGATGACCCCTGCGAGGACCTCCAGCTGAGACAAACCCAGGCCCGACAACTCAGGC 360  
QY 361 CCGAACAACCAACGAGGCCCGCAACCACTACTAGAGCCCTTACTCAACCAACCCAGGC 420  
| | | | |  
Db 361 CCGAACAACCAACGAGGCCCGCAACCACTACTAGAGCCCTTACTCAACCAACCCAGGC 420  
QY 421 CCCAACCAACCAACGAGGCCCGCAACCACTACTAGAGCCCGCAACCAACCCAGGC 480  
| | | | |  
Db 421 CCGAACAACCAACGAGGCCCGCAACCACTACTAGAGCCCGCAACCAACCCAGGC 480  
QY 481 AACTACCACTAGAGCCCTTACTACTACTAGAGCCCGCAACCACTAGAGCCCGC 540  
| | | | |  
Db 481 AACTACCACTAGAGCCCTTACTACTACTAGAGCCCGCAACCACTAGAGCCCGC 540  
QY 541 TACCACCAACCAACGAGGCCCGCAACCACTACTAGAGCCCGCAACCACTAGAGCCCGC 600  
| | | | |  
Db 541 TACCACCAACCAACGAGGCCCGCAACCACTACTAGAGCCCGCAACCACTAGAGCCCGC 600  
QY 601 TACCACCAACCAACGAGGCCCGCAACCACTACTAGAGCCCGCAACCACTAGAGCCCGC 660  
| | | | |  
Db 601 TACCACCAACCAACGAGGCCCGCAACCACTACTAGAGCCCGCAACCACTAGAGCCCGC 660  
QY 661 GGCAGCACTAGAGCCCGCAACCACTACTAGAGCCCGCAACCACTAGAGCCCGC 720  
| | | | |  
Db 661 GGCAGCACTAGAGCCCGCAACCACTACTAGAGCCCGCAACCACTAGAGCCCGC 720  
QY 721 AGGTGTTTCTGCAACCACTACTAGAGCCCGCAACCACTACTAGAGCCCGC 780  
| | | | |  
Db 721 AGGTGTTTCTGCAACCACTACTAGAGCCCGCAACCACTACTAGAGCCCGC 780  
QY 781 TGGTTGGCCAGCTGAGCTTCCAGCATCTACTGTTGATTCGCCAGCAAGTACTCAACT 840  
| | | | |  
Db 781 TGGTTGGCCAGCTGAGCTTCCAGCATCTACTGTTGATTCGCCAGCAAGTACTCAACT 840  
QY 841 CTCTCTACCACTGCTCCAGAGGTTACACTTTCGAAACAGAGGTCCTGAGGAGCTACTT 900  
| | | | |  
Db 841 CTCTCTACCACTGCTCCAGAGGTTACACTTTCGAAACAGAGGTCCTGAGGAGCTACTT 900  
QY 901 CAACCCCTAGCTCCAGGCTGCGGCTCTCTGCTAACTGTTGAATGGCAGGCGGAAACAG 960  
| | | | |  
Db 901 CAACCCCTAGCTCCAGGCTGCGGCTCTCTGCTAACTGTTGAATGGCAGGCGGAAACAG 960  
QY 961 CCCCAGCCCGCAGCTCAGAGAGCAAGAGAGAGATTCGATTCGAGAGCTCTCT 1020  
| | | | |  
Db 961 CCCCAGCCCGCAGCTCAGAGAGCAAGAGAGAGATTCGATTCGAGAGCTCTCT 1020  
QY 1021 CGACATGATGCTCCAGCTTGAATTCGAGTGGCTCTTGGCCGACGAGAAACGCTTG 1080  
| | | | |  
Db 1021 CGACATGATGCTCCAGCTTGAATTCGAGTGGCTCTTGGCCGACGAGAAACGCTTG 1080  
QY 1081 CGACATGATGCTCCAGCTTGAATTCGAGTGGCTCTTGGCCGACGAGAAACGCTTG 1140  
| | | | |  
Db 1081 CGACATGATGCTCCAGCTTGAATTCGAGTGGCTCTTGGCCGACGAGAAACGCTTG 1140  
QY 1141 CCACCTCAGTTTGAATTCGAGTGGCTCTTGGCCGACGAGAAACGCTTG 1200  
| | | | |  
Db 1141 CCACCTCAGTTTGAATTCGAGTGGCTCTTGGCCGACGAGAAACGCTTG 1200  
QY 1201 CCGCGGCGAGAGCGAAGTGAAGTTCGAGAGAGATGCTTCGACCGGCTGTTACTGCC 1260  
| | | | |  
Db 1201 CCGCGGCGAGAGCGAAGTGAAGTTCGAGAGAGATGCTTCGACCGGCTGTTACTGCC 1260  
QY 1261 CAGGGAACCATTTGAATTCGAGAGAGATGCTTCGAGAGAGATGCTTCGAGAGAGAT 1320  
| | | | |  
Db 1261 CAGGGAACCATTTGAATTCGAGAGAGATGCTTCGAGAGAGATGCTTCGAGAGAGAT 1320  
QY 1321 CCACCTCCTCCCGCAG 1380  
| | | | |  
Db 1321 CCACCTCCTCCCGCAG 1380  
QY 1381 TATCGCAAGACCTTGGCCCTGAGAAACCTGACATTCAGTCTGCGCACAGAGTCTGTGAGTC 1440

Db 1381 TATCGCAAGACCTTGGCCCTGAGAAACCTGACATTCAGTCTGCGCACAGAGTCTGTGAGTC 1440  
| | | | |  
QY 1441 TCGTGAGCCGCTGTTGGCCAACTTTTCAGAGTGCATTCGAGAGAGAGAGAGAGAGAG 1500  
| | | | |  
Db 1441 TCGTGAGCCGCTGTTGGCCAACTTTTCAGAGTGCATTCGAGAGAGAGAGAGAGAGAG 1500  
QY 1501 TCGTGAGCCGCTGTTGGCCAACTTTTCAGAGTGCATTCGAGAGAGAGAGAGAGAGAG 1560  
| | | | |  
Db 1501 TCGTGAGCCGCTGTTGGCCAACTTTTCAGAGTGCATTCGAGAGAGAGAGAGAGAGAG 1560  
QY 1561 TCGCGACCCCTCCAGCTGTTCCAGCTGACGACGACCCGCACTGAGAGAGAGAGAGAG 1620  
| | | | |  
Db 1561 TCGCGACCCCTCCAGCTGTTCCAGCTGACGACGACCCGCACTGAGAGAGAGAGAGAG 1620  
QY 1621 ACCTACCAACCGCAATTCATCTCGGCGCCCGCAACGAGTGCAGAGAGAGAGAGAGAG 1680  
| | | | |  
Db 1621 ACCTACCAACCGCAATTCATCTCGGCGCCCGCAACGAGTGCAGAGAGAGAGAGAGAG 1680  
QY 1681 TCGTGAGCCGCTGTTGGCCAACTTTTCAGAGTGCATTCGAGAGAGAGAGAGAGAGAG 1740  
| | | | |  
Db 1681 TCGTGAGCCGCTGTTGGCCAACTTTTCAGAGTGCATTCGAGAGAGAGAGAGAGAGAG 1740  
QY 1741 CCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
| | | | |  
Db 1741 CCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1800  
QY 1801 CACGCGTGGCCCGCAG 1860  
| | | | |  
Db 1801 CACGCGTGGCCCGCAG 1860  
QY 1861 AACTAGTGTACATAC 1920  
| | | | |  
Db 1861 AACTAGTGTACATAC 1920  
QY 1921 CACAGTACTGTAC 1980  
| | | | |  
Db 1921 CACAGTACTGTAC 1980  
QY 1981 AGTCAGTGTAC 2040  
| | | | |  
Db 1981 AGTCAGTGTAC 2040  
QY 2041 CAAC 2100  
| | | | |  
Db 2041 CAAC 2100  
QY 2101 TGAATTCGAG 2160  
| | | | |  
Db 2101 TGAATTCGAG 2160  
QY 2161 TTTGGCCCTGAGGCTCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
| | | | |  
Db 2161 TTTGGCCCTGAGGCTCAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2220  
QY 2221 CCGTGAGTGTGCTGAGAGGCTCCAGTTCACACCCCACTACTAAGACCTGTGACTTCGC 2280  
| | | | |  
Db 2221 CCGTGAGTGTGCTGAGAGGCTCCAGTTCACACCCCACTACTAAGACCTGTGACTTCGC 2280  
QY 2281 TTTGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340  
| | | | |  
Db 2281 TTTGCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340  
QY 2341 GGTCTTCATCCCATGGAACAACTAATGAGAGATCAGAGAGAGAGAGAGAGAGAGAG 2400  
| | | | |  
Db 2341 GGTCTTCATCCCATGGAACAACTAATGAGAGATCAGAGAGAGAGAGAGAGAGAGAG 2400  
QY 2401 GTAAACCTACTTAATTAATGAAGGTTTGTGTTTAAAAAATGAGAGAGAGAGAGAGAG 2455  
| | | | |  
Db 2401 GTAAACCTACTTAATTAATGAAGGTTTGTGTTTAAAAAATGAGAGAGAGAGAGAGAG 2455

US-09-294-663-1  
Sequence 1, Application US/09294663.  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
APPLICANT: Wang, Ping  
TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
TITLE OF INVENTION: cdna and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pimisti & Michaels, P.C.  
STREET: 118 North Tiooga Street  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/294,663  
FILING DATE: 19-APR-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/103,429  
FILING DATE: 24-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ. ID NO.: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2455 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: Peritrophic Membrane  
IMMEDIATE SOURCE:  
CLONE: IM14  
US-09-294-663-1

Query Match 100.0%; Score 2455; DB 38; Length 2455;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GRAAGCTTAAGTGAAGATACGACGCAAGTATGATTAAGACCTCTATTCT 60  
DB 1 GTRACCTTAAGTGAAGATTAACGACGCAAGTATGATTAAGACCTCTATTCT 60  
QY 61 GAGGCGCTCGGCGCTGTCGCGCGGCTCTGAAGTCAGGAGCGGAGAAACCCGC 120  
DB 61 GAGGCGCTCGGCGCTGTCGCGCGGCTCTGAAGTCAGGAGCGGAGAAACCCGC 120  
QY 121 TCTCCAGGAGCCGACCCAGACTGCCCTCCGCTGAGCAGCACTGCTCTCACA 180  
DB 121 TCTCCAGGAGCCGACCCAGACTGCCCTCCGCTGAGCAGCACTGCTCTCACA 180  
QY 181 ATAGGAGCTGACCAAGTCTACTACTGTAATATGATTCGAATTCAGCAGGAGA 240  
DB 181 ATAGGAGCTGACCAAGTCTACTACTGTAATATGATTCGAATTCAGCAGGAGA 240  
QY 241 CTGTGCTCTGTACCAATTCAAGTTCTCCGCTCAGACTTGTTCAGCGCCTTAGC 300

DB 241 CTGTGCTCTGTACCAATTCAAGTTCTCCGCTCAGACTTGTTCAGCGCCTTAGC 300  
QY 301 CGGATGACACCTGCGCAGGAGCTGCGAGTGAAGAACCCAGGCGCCAGAACTCAGGC 360  
DB 301 CGGATGACACCTGCGCAGGAGCTGCGAGTGAAGAACCCAGGCGCCAGAACTCAGGC 360  
QY 361 CCCAACCAACCCAGGCGCCCAACCACTACTAGGCGCTTACTACACCCAGGC 420  
DB 361 CCCAACCAACCCAGGCGCCCAACCACTACTAGGCGCTTACTACACCCAGGC 420  
QY 421 CCCAACCAACCCAGGCGCCCAACCACTACTAGGCGCTTACTACACCCAGGC 480  
DB 421 CCCAACCAACCCAGGCGCCCAACCACTACTAGGCGCTTACTACACCCAGGC 480  
QY 481 AACTGACCTGAGGCGCTTACTAGCACTGAGGCGCCCAACCACTCAGGCGCC 540  
DB 481 AACTGACCTGAGGCGCTTACTAGCACTGAGGCGCCCAACCACTCAGGCGCC 540  
QY 541 TACCAACCAACCCAGGCGCCCAACCACTCAGGCGCCCAACCACTCAGGCGCC 600  
DB 541 TACCAACCAACCCAGGCGCCCAACCACTCAGGCGCCCAACCACTCAGGCGCC 600  
QY 601 TACCACTAGGCGCCCAACTACCAATCAGGCGCCCAACTAGGCGCCCAACTAGCC 660  
DB 601 TACCACTAGGCGCCCAACTACCAATCAGGCGCCCAACTAGGCGCCCAACTAGCC 660  
QY 661 GCGCGCAATACCCGCGCGCACTACCCCTGCGGAGCAACCCCGCTGCACTACCC 720  
DB 661 GCGCGCAATACCCGCGCGCACTACCCCTGCGGAGCAACCCCGCTGCACTACCC 720  
QY 721 AGGTGCTCTGCAACCACTTACGCGCGGAGTGGCGCGCGCTGTAAGCTTGGCCAA 780  
DB 721 AGGTGCTCTGCAACCACTTACGCGCGGAGTGGCGCGCGCTGTAAGCTTGGCCAA 780  
QY 781 TGGTGGCCGAGGTGACCTTGCACATCCACTTGTGATTCGCCAGCAAGTACTGCACT 840  
DB 781 TGGTGGCCGAGGTGACCTTGCACATCCACTTGTGATTCGCCAGCAAGTACTGCACT 840  
QY 841 CTTCACGAGTCTGCAAGGTTACACCTTGAACAGAGAGTCCCTGAGGACTACTT 900  
DB 841 CTTCACGAGTCTGCAAGGTTACACCTTGAACAGAGAGTCCCTGAGGACTACTT 900  
QY 901 CAACCCCTAGCTCAGCGCTGCGACTCTCTGTAAGTGAAGTGAAGGAGGAAATCAG 960  
DB 901 CAACCCCTAGCTCAGCGCTGCGACTCTCTGTAAGTGAAGTGAAGGAGGAAATCAG 960  
QY 961 CCGCGCAGCCGAGTCAAGAGGCAAGCAAGCAAGCAATGATGATGAGACTCTCT 1020  
DB 961 CCGCGCAGCCGAGTCAAGAGGCAAGCAAGCAAGCAATGATGATGAGACTCTCT 1020  
QY 1021 CGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
DB 1021 CGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1080  
QY 1081 CGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
DB 1081 CGACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1140  
QY 1141 CCACCTCACTTTTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1200  
DB 1141 CCACCTCACTTTTGAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCAAGTCA 1200  
QY 1201 CGGCGCGAGAGCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1260  
DB 1201 CGGCGCGAGAGCGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 1260  
QY 1261 CAGGGAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
DB 1261 CAGGGAACCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1320  
QY 1321 CCACT 1380  
DB 1321 CCACT 1380

Db 1321 CCACTCTCTCCCCAGAGAGCGAGCTGGCCAGTATCTACAGTGTGTCCATGAGACAGAC 1380  
Qy 1381 TATGCAAGAGACTTGGCCCTGGAAACCTGCACTTTCAGTCTGCGCACAGAGTCTGTGAGTC 1440  
Db 1381 TATGCAAGAGACTTGGCCCTGGAAACCTGCACTTTCAGTCTGCGCACAGAGTCTGTGAGTC 1440  
Qy 1441 TCTGTGACGCGTGTGTTGCCAAGTTTTCGAGTGGGATTTTGACACCAAGTGCATGTGAC 1500  
Db 1441 TCTGTGACGCGTGTGTTGCCAAGTTTTCGAGTGGGATTTTGACACCAAGTGCATGTGAC 1500  
Qy 1501 TGTGCCCCGACAGCTCTCCAAAGGCTGGCCCCCAAGGGCTGCCCAAGGGTGGCCCCAAC 1560  
Db 1501 TGTGCCCCGACAGCTCTCCAAAGGCTGGCCCCCAAGGGCTGCCCAAGGGTGGCCCCAAC 1560  
Qy 1561 TGGCGACCCCTCCAGCCTGTGTCCACCTGCAAGCAGCCGCAACTGCAGCCCAAGTCCC 1620  
Db 1561 TGGCGACCCCTCCAGCCTGTGTCCACCTGCAAGCAGCCGCAACTGCAGCCCAAGTCCC 1620  
Qy 1621 ACCTACACCGCAATTCTCTACTCCGGCCCCCACCCTGCCCCCAGCGAGCTCTACTAC 1680  
Db 1621 ACCTACACCGCAATTCTCTACTCCGGCCCCCACCCTGCCCCCAGCGAGCTCTACTAC 1680  
Qy 1681 TGTGCCCCGACAGCTCTCCAAAGGCTGGCCCCCAAGGGCTGCCCAAGGGTGGCCCCAAC 1740  
Db 1681 TGTGCCCCGACAGCTCTCCAAAGGCTGGCCCCCAAGGGCTGCCCAAGGGTGGCCCCAAC 1740  
Qy 1741 CCCTACTACTGTGTCTCCCTGAATCCCAATCAGTGTACATCAGAGCGCTACCGCTGCCCC 1800  
Db 1741 CCCTACTACTGTGTCTCCCTGAATCCCAATCAGTGTACATCAGAGCGCTACCGCTGCCCC 1800  
Qy 1801 CACCGCTGCCCCCAGCGCTGCCCCCAGCGAGCCCTACTACTGTGTGTCCAGAAATCCC 1860  
Db 1801 CACCGCTGCCCCCAGCGCTGCCCCCAGCGAGCCCTACTACTGTGTGTCCAGAAATCCC 1860  
Qy 1861 AACTACTGTACATCAGT 1920  
Db 1861 AACTACTGTACATCAGT 1920  
Qy 1921 CACAGTCTACTGTACACCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980  
Db 1921 CACAGTCTACTGTACACCACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1980  
Qy 1981 AGTACTGTGTACACCACTGT 2040  
Db 1981 AGTACTGTGTACACCACTGT 2040  
Qy 2041 CACACCCAGAGCTGCCCCGGGTAATACACAGGCGACAGCTGTGTGTGTGTGTGTGTGTGT 2100  
Db 2041 CACACCCAGAGCTGCCCCGGGTAATACACAGGCGACAGCTGTGTGTGTGTGTGTGTGTGT 2100  
Qy 2101 TGACATGAGACCCCTCTCCCAAGAGACCCCATACACCTTGTGTGTGTGTGTGTGTGTGT 2160  
Db 2101 TGACATGAGACCCCTCTCCCAAGAGACCCCATACACCTTGTGTGTGTGTGTGTGTGTGT 2160  
Qy 2161 TTTGGCCCTGGGCTGACGCTGT 2220  
Db 2161 TTTGGCCCTGGGCTGACGCTGT 2220  
Qy 2221 CCGT 2280  
Db 2221 CCGT 2280  
Qy 2281 TTTGCAACGCTGT 2340  
Db 2281 TTTGCAACGCTGT 2340  
Qy 2341 GGTCTTATCCATGAGAACTAGATGAAGACATCAGAGCGCTGTGAATCTTGTGAGTT 2400  
Db 2341 GGTCTTATCCATGAGAACTAGATGAAGACATCAGAGCGCTGTGAATCTTGTGAGTT 2400  
Qy 2401 GTAACCTACTTAATTAATGAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2455  
Db 2401 GTAACCTACTTAATTAATGAAGTTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 2455

RESULT 4  
US-09-103-429-2  
; Sequence 2 Application US/09103429  
; GENERAL INFORMATION:  
; APPLICANT: Granados, Robert R  
; APPLICANT: Wang, Ping  
; TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
; TITLE OF INVENTION: CDNA and Related Products and Methods  
; NUMBER OF SEQUENCES: 2  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Brown, P. C.  
; STREET: 118 North Tloga  
; CITY: Ithaca  
; STATE: NY  
; COUNTRY: USA  
; ZIP: 14850  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/09/103,429  
; FILING DATE: 24-JUN-1998  
; CLASSIFICATION: 800  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Michaels, Christopher A  
; REGISTRATION NUMBER: 34,390  
; REFERENCE/DOCKET NUMBER: BTI-39  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (607) 256-2000  
; TELEFAX: (607) 256-3628  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 2821 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FRAGMENT TYPE: N-terminal  
; ORIGINAL SOURCE:  
; ORGANISM: Trichoplusia ni  
; DEVELOPMENTAL STAGE: larva  
; TISSUE TYPE: peritrophic membrane  
; US-09-103-429-2  
Query Match 95.9%; Score 2354.4; DB 33; Length 2821;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;  
Qy 13 GAAAGAAATTAACGAGCAACAACTATGATTAAGACCTCTCTTCTGAGCGCCCTCG 72  
Db 1 GAAAGAAATTAACGAGCAACAACTATGATTAAGACCTCTCTTCTGAGCGCCCTCG 60  
Qy 73 GCTGCTGCGCGCGCTCTGAGTCAAGCAGCGGAGAGAACCCGCTCTCCAGAGCC 132  
Db 61 GCTGCTGCGCGCGCTCTGAGTCAAGCAGCGGAGAGAACCCGCTCTCCAGAGCC 120  
Qy 133 GCACCCAGAGTGGCCGCCGCTGAGAGAGAGCTGCTGCTGCTGAGATAGAGCTGCAC 192  
Db 121 GCACCCAGAGTGGCCGCCGCTGAGAGAGAGCTGCTGCTGCTGAGATAGAGCTGCAC 180  
Qy 193 CAAGTTCTACTGTGAATATGCTCAAGTTCATGCGACCGAGAGAGTGTCTCTCTG 252  
Db 181 CAAGTTCTACTGTGAATATGCTCAAGTTCATGCGACCGAGAGAGTGTCTCTCTG 240  
Qy 253 TACCGAATTCAGTTCTCCGCTCAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 312  
Db 253 TACCGAATTCAGTTCTCCGCTCAGACTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 312

Db 241 TACCAATTCAAGTTCCTCCGCTCAGACTGTGTTCAGCGCCCTTTAGCCGGATGCACCT 300  
Oy 313 GCCAGAGACCTCCAGCTGAGACAACCCAGGCCAGCAGAACACTAGAGGCCCAACAACAC 372  
Db 301 GCCAGAGACCTCCAGCTGAGACAACCCAGGCCAGCAGAACACTAGAGGCCCAACAACAC 360  
Oy 373 CCAGGCCCAACCACTACTCAGGCCCTTACTTACACACCCAGGCCCAACAACAC 432  
Db 361 CCAGGCCCAACCACTACTCAGGCCCTTACTTACACACCCAGGCCCAACAACAC 420  
Oy 433 CAGCCAGGCCCAACCACTAGGCCCTTACACACCCAGGCCCTTACACACCTA 492  
Db 421 CAGCCAGGCCCAACCACTAGGCCCTTACACACCCAGGCCCTTACACACCTA 480  
Oy 493 GAGCCCTACTACTACTCAGGCCCTTACACACCCAGGCCCTTACACACCTA 552  
Db 481 GAGCCCTACTACTACTCAGGCCCTTACACACCCAGGCCCTTACACACCTA 540  
Oy 553 CAGGCCCAACCACTAGGCCCTTACACACCCAGGCCCTTACACACCTA 612  
Db 541 CAGGCCCAACCACTAGGCCCTTACACACCCAGGCCCTTACACACCTA 600  
Oy 613 CCAACTACAACTACAGGCCCTTACACACCCAGGCCCTTACACACCTA 672  
Db 601 CCAACTACAACTACAGGCCCTTACACACCCAGGCCCTTACACACCTA 660  
Oy 673 CCGGCGCGCACTACCCCTGCGCGCAACCCCGCTGCACTACCCAGGTTCTCTGC 732  
Db 661 CCGGCGCGCACTACCCCTGCGCGCAACCCCGCTGCACTACCCAGGTTCTCTGC 720  
Oy 733 ACCCACTTACGCCCAAGTCTGCGCGCAACCCCGCTGCACTACCCAGGTTCTCTGC 792  
Db 721 ACCCACTTACGCCCAAGTCTGCGCGCAACCCCGCTGCACTACCCAGGTTCTCTGC 780  
Oy 793 TACCTTGCACATCCACTTGTGATTCGCCAGCAAGTACTGCAACCTTCTTACAGAG 852  
Db 781 TACCTTGCACATCCACTTGTGATTCGCCAGCAAGTACTGCAACCTTCTTACAGAG 840  
Oy 853 CTCACAGGTTACACTTGTGAGAGAGTGCCTGAGGAGCTTACTTACCCCTAGCT 912  
Db 841 CTCACAGGTTACACTTGTGAGAGAGTGCCTGAGGAGCTTACTTACCCCTAGCT 900  
Oy 913 CCAGGCGTGCAGCTCTCTGCTAACGTTGAATGCAGGCGGAATACGCCCGCAACCCC 972  
Db 901 CCAGGCGTGCAGCTCTCTGCTAACGTTGAATGCAGGCGGAATACGCCCGCAACCCC 960  
Oy 973 ACTCACAAGAGCAAGAGAGAGAGACATTGACATCGAGACCTCTCTGACAAATGATG 1032  
Db 961 ACTCACAAGAGCAAGAGAGAGAGACATTGACATCGAGACCTCTCTGACAAATGATG 1020  
Oy 1033 CCAGCTAACCTTGAAGTGAAGTGGCTTGGCCCAAGGAAACCGTTGAGACAATTA 1092  
Db 1021 CCAGCTAACCTTGAAGTGAAGTGGCTTGGCCCAAGGAAACCGTTGAGACAATTA 1080  
Oy 1093 CCAAGTGCCTCAGAGTGAAGTGGTGAAGAGGCTTGGAGGCGGACCCACTTCACTT 1152  
Db 1081 CCAAGTGCCTCAGAGTGAAGTGGTGAAGAGGCTTGGAGGCGGACCCACTTCACTT 1140  
Oy 1153 TGAATCTCAGCAATGTGACCAATCGAGTCTGTGCTCACCCTCCCGGCGGAGAG 1212  
Db 1141 TGAATCTCAGCAATGTGACCAATCGAGTCTGTGCTCACCCTCCCGGCGGAGAG 1200  
Oy 1213 CGAAGAGTTGAGAGTGCAGAGAGATCTGACCCGCTGGTACTGCGCCAGGAACCAT 1272  
Db 1201 CGAAGAGTTGAGAGTGCAGAGAGATCTGACCCGCTGGTACTGCGCCAGGAACCAT 1260  
Oy 1273 TGAATGGAGCCCTCCCAAGGCTGCCCTGCGCACTTCAAGATGACACCTCTCTCC 1332  
Db 1261 TGAATGGAGCCCTCCCAAGGCTGCCCTGCGCACTTCAAGATGACACCTCTCTCC 1320  
Oy 1333 CCACGAGACGAGTGGCGGCAATATACAGTGTGTCCATGGACAGATATGCAAGACC 1392  
Db 1321 CCACGAGACGAGTGGCGGCAATATACAGTGTGTCCATGGACAGATATGCAAGACC 1380

Oy 1393 TTGCCCTGGAAAACCTGCACTTCACTGCTGCGACACAGTCTGTGAGTCTCTGTGACCG 1452  
Db 1381 TTGCCCTGGAAAACCTGCACTTCACTGCTGCGACACAGTCTGTGAGTCTCTGTGACCG 1440  
Oy 1453 TGTGTTGCCAAGTTTTCGAGTGTGATTTGACAAACAGTGCATGTGATGCTGCTGCCAC 1512  
Db 1441 TGTGTTGCCAAGTTTTCGAGTGTGATTTGACAAACAGTGCATGTGATGCTGCTGCCAC 1500  
Oy 1513 AGCTGCTCCAAAGGCTGCGCCCAAGGCGGCGCAAGGCGTGGCCCACTGCGCGACCTTC 1572  
Db 1501 AGCTGCTCCAAAGGCTGCGCCCAAGGCGGCGCAAGGCGTGGCCCACTGCGCGACCTTC 1560  
Oy 1573 CAGCGTGTCCCACTTGCAGCGCACCGGCAACTGACAGCCCACTGCGCGCACTGACACCG 1632  
Db 1561 CAGCGTGTCCCACTTGCAGCGCACCGGCAACTGACAGCCCACTGCGCGCACTGACACCG 1620  
Oy 1633 AATTCTTACTCGGCGCCCAAGCGCTGCGCCCAAGCGAGCTCTTACTTACTGCTGCCCTGA 1692  
Db 1621 AATTCTTACTCGGCGCCCAAGCGCTGCGCCCAAGCGAGCTCTTACTTACTGCTGCCCTGA 1680  
Oy 1693 ATCCCAACCACTGTCAGAGTACACAGTACAGTACGCTGCTGCCAAGCGAGCGCTTACTGTC 1752  
Db 1681 ATCCCAACCACTGTCAGAGTACACAGTACAGTACGCTGCTGCCAAGCGAGCGCTTACTGTC 1740  
Oy 1753 TGTCCCTGAATCCCAATCCACTGTCATCAGCGCTACCGCTGCGCCCAAGCGCTGCGCC 1812  
Db 1741 TGTCCCTGAATCCCAATCCACTGTCATCAGCGCTACCGCTGCGCCCAAGCGCTGCGCC 1800  
Oy 1813 CAGCGCTGCGCCCAAGCGCGAGCGCTTACTGCTGCTGCTGCCAAGAAATCCCACTTGTGAC 1872  
Db 1801 CAGCGCTGCGCCCAAGCGCGAGCGCTTACTGCTGCTGCTGCCAAGAAATCCCACTTGTGAC 1860  
Oy 1873 ATCACCACCTTACTGCTGCGCCCAAGCGAGCACTGCTGCCAAGAAATCCCACTTGTGAC 1932  
Db 1861 ATCACCACCTTACTGCTGCGCCCAAGCGAGCACTGCTGCCAAGAAATCCCACTTGTGAC 1920  
Oy 1933 ACCACCACTGCT----- 1946  
Db 1921 ACCACCACTGCTGCTGCTTACTACCGGACAGCACTGCTGCCAAGAAATCCCACTTGTGAC 1980  
Oy 1946 -----GCCCTTACTACCGGACAGCACTGCTGCCAAGAAATCCCACTTGTGAC 1995  
Db 1981 ACCACCGAGCGCTTACTACCGGAGCACTGCTGCCAAGAAATCCCACTTGTGAC 2040  
Oy 1996 CACTGCTGCGCCCAAGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2055  
Db 2041 CACTGCTGCGCCCAAGCGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2100  
Oy 2056 CCGGTACTACCAACAGGCGAGCACTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2115  
Db 2101 CCGGTACTACCAACAGGCGAGCACTACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2160  
Oy 2116 TCTCCCAAGCAAGCCCACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2175  
Db 2161 TCTCCCAAGCAAGCCCACTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2220  
Oy 2176 CGCTGACTGCGCAAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2235  
Db 2221 CGCTGACTGCGCAAAATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2280  
Oy 2236 TGAAGGTCTCCGTTCAACCCCACTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2295  
Db 2281 TGAAGGTCTCCGTTCAACCCCACTTGAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2340  
Oy 2296 CGTGAAGAGCAATTCAGATGTCTGAAGGTACGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 2355  
Db 2341 CGTGAAGAGCAATTCAGATGTCTGAAGGTACGAAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 2400  
Oy 2356 GACAAATCTAGTGAAGATCAGACAGGCGTGAATTTGATTTGATTTGATTTGATTTGATTTGATTT 2415  
Db 2401 GACAAATCTAGTGAAGATCAGACAGGCGTGAATTTGATTTGATTTGATTTGATTTGATTTGATTT 2460



QY 2416 TTAATGAGCTTTGTTT 2435  
DB 2461 TTAATGAGCTTTGTTT 2480

## RESULT 5

US-09-103-429A-2

Sequence 2, Application US/09103429A

GENERAL INFORMATION:

APPLICANT: Granados, Robert R

APPLICANT: Many, Ping

TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin

TITLE OF INVENTION: CDNA and Related Products and Methods

NUMBER OF SEQUENCES: 4

CORRESPONDENCE ADDRESS:

ADDRESSEE: Brown, Pinisi &amp; Michaels, P.C.

STREET: 118 North Tloga

CITY: Ithaca

STATE: NY

COUNTRY: USA

ZIP: 14850

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/103,429A

FILING DATE: 24-JUN-1998

CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:

NAME: Michaels, Christopher A

REGISTRATION NUMBER: 34,390

REFERENCE/DOCKET NUMBER: BTI-39

TELECOMMUNICATION INFORMATION:

TELEPHONE: (607) 256-2000

TELEFAX: (607) 256-3628

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 2821 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: not relevant

MOLECULE TYPE: cDNA

HYPOTHETICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal

ORIGINAL SOURCE:

ORGANISM: Trichoplusia ni

DEVELOPMENTAL STAGE: larva

TISSUE TYPE: peritrophic membrane

US-09-103-429A-2

Query Match 95.9%; Score 2354.4; DB 33; Length 2821;

Best Local Similarity 97.7%; Pred. No. 0;

Matches 2422; Conservative 0; Mismatches 1; Indels 57; Gaps 1;

QY 13 GAAAGAAATACAGGAGCAAGTATGATTAAGACCTCTCTATCTCTGAGGCGCTCG 72  
DB 1 GAAAGAAATACAGGAGCAAGTATGATTAAGACCTCTCTATCTCTGAGGCGCTCG 60  
QY 73 GCTGCTGCGCGCGCTCTGAGTCAAGTCAAGCGAGGAGAGAACCCGCTCTCAAGAGCC 132  
DB 61 GCTGCTGCGCGCGCTCTGAGTCAAGTCAAGCGAGGAGAGAACCCGCTCTCAAGAGCC 120  
QY 133 GCACCCAGACTGCGCTCTGAGTCAAGTCAAGCGAGGAGAGAACCCGCTCTCAAGAGCC 192  
DB 121 GCACCCAGACTGCGCTCTGAGTCAAGTCAAGCGAGGAGAGAACCCGCTCTCAAGAGCC 180  
QY 193 CAAGTTCTACTAGTGAATATGATGTCAGAGTTCATGCGACCGAGAGACGTGCTCTG 252  
DB 181 CAAGTTCTACTAGTGAATATGATGTCAGAGTTCATGCGACCGAGAGACGTGCTCTG 240

QY 253 TACCGAATTCAGATTCTCCGCTCAGACTGTGTTCAGCGCGCTTTAGCCGATGCAACCT 312  
DB 241 TACCGAATTCAGATTCTCCGCTCAGACTGTGTTCAGCGCGCTTTAGCCGATGCAACCT 300  
QY 313 GCGAGAGCTCCAGCTGAGACAAACCCAGGCGCCAGAGAACACTCAGGCGCCCAACACAC 372  
DB 301 GCGAGAGCTCCAGCTGAGACAAACCCAGGCGCCAGAGAACACTCAGGCGCCCAACACAC 360  
QY 373 CGAGGCGCCCAACCACTACTAGGCGCCCTACTACAAACCCAGGCGCCCAACCAACAC 432  
DB 361 CGAGGCGCCCAACCACTACTAGGCGCCCTACTACAAACCCAGGCGCCCAACCAACAC 420  
QY 433 CACCGAGGCGCCCAACCACTACTAGGCGCCCTACTACAAACCCAGGCGCCCAACCACT 492  
DB 421 CACCGAGGCGCCCAACCACTACTAGGCGCCCTACTACAAACCCAGGCGCCCAACCACT 480  
QY 493 GCGGCTTACTACTAGTCAAGTCAAGGCGCCCAACCACTACTAGGCGCCCTACTACAAAC 552  
DB 481 GCGGCTTACTACTAGTCAAGTCAAGGCGCCCAACCACTACTAGGCGCCCTACTACAAAC 540  
QY 553 CGAGGCGCCCAACCACTACTAGGCGCCCTACTACAAACCCAGGCGCCCAACCACTAGG 612  
DB 541 CGAGGCGCCCAACCACTACTAGGCGCCCTACTACAAACCCAGGCGCCCAACCACTAGG 600  
QY 613 CCGAATTAATCAACCAAGGCTGCAACTACCCGCGCACTACCCGCGCGCAACTAC 672  
DB 601 CCGAATTAATCAACCAAGGCTGCAACTACCCGCGCGCACTACCCGCGCGCAACTAC 660  
QY 673 CCGGCGCGCAACTACCCGCGCGCGCAACTACCCGCGCGCAACTACCCGCGCGCAACT 732  
DB 661 CCGGCGCGCAACTACCCGCGCGCGCAACTACCCGCGCGCAACTACCCGCGCGCAACT 720  
QY 733 ACCGCTTACG 792  
DB 721 ACCGCTTACG 780  
QY 793 TGACTTGCAGATCACTTGTGATTCGCGAGCAAGTACTGCAACTCTTCTTACAGTG 852  
DB 781 TGACTTGCAGATCACTTGTGATTCGCGAGCAAGTACTGCAACTCTTCTTACAGTG 840  
QY 853 CTCGACAGGTTACCTTGGAGACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912  
DB 841 CTCGACAGGTTACCTTGGAGACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
QY 913 CGAGGCG 972  
DB 901 CGAGGCG 960  
QY 973 AGTCACAGAGGCGAGAGAGAGAGAGAGATGACATCGAGACCTCTCGACAAATGATG 1032  
DB 961 AGTCACAGAGGCGAGAGAGAGAGAGAGATGACATCGAGACCTCTCGACAAATGATG 1020  
QY 1033 CCGAGTAACTTGAATGAGTGTGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1092  
DB 1021 CCGAGTAACTTGAATGAGTGTGCTTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1080  
QY 1093 CGAGTGTGCGAGTGTGATGAGAGGCGGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 1152  
DB 1081 CGAGTGTGCGAGTGTGATGAGAGGCGGTGTGAGCGCGCGCGCGCGCGCGCGCGCGCG 1140  
QY 1153 TGAATTCAGCAATGAGACATCGAGTGTGCTGCGACCCCTCCGCGCGCGAGAG 1212  
DB 1141 TGAATTCAGCAATGAGACATCGAGTGTGCTGCGACCCCTCCGCGCGCGAGAG 1200  
QY 1213 CGAAGAGTTGAGTGTGAGAGAGATGCTGACCGCGCTGTACTGCGCGCGAGACCAT 1272  
DB 1201 CGAAGAGTTGAGTGTGAGAGAGATGCTGACCGCGCTGTACTGCGCGCGAGACCAT 1260  
QY 1273 TGAATGAGAGCG 1332  
DB 1261 TGAATGAGAGCG 1320





|||||  
Db 121 GCACCCAGATGCTCCCTCCCTGAGCAGCAGCTGCTCTCTCCACGAATACAGCTGCAC 180  
QY 193 CAAGTTCTACTGTAATATGATGTCGAAGTTCAATGCGACGAGAGACTGTGCTCTGG 252  
Db 181 CAGATTCTACTGTAATATGATGTCGAAGTTCAATGCGACGAGAGACTGTGCTCTGG 240  
QY 253 TACCGAATTCAGATTCTCCGCTGAGACTGTTGTTCAAGCGGCTTTAGCCGGATGACCTT 312  
Db 241 TACCGAATTCAGATTCTCCGCTGAGACTGTTGTTCAAGCGGCTTTAGCCGGATGACCTT 300  
QY 313 GCCAGAGACTCCAGCTGAGAGAACCCAGGCCCCAGCAACAATCAGGCCCCAACACAC 372  
Db 301 GCCAGAGACTCCAGCTGAGAGAACCCAGGCCCCAGCAACAATCAGGCCCCAACACAC 360  
QY 373 CCAGGCCCCAACCACTCTAGGCCCCCTACTACAAACCAACGAGCCCCAACCAAC 432  
Db 361 CCAGGCCCCAACCACTCTAGGCCCCCTACTACAAACCAACGAGCCCCAACCAAC 420  
QY 433 CACCCAGGCCCCAACCACTCTAGGCCCCAACCAACGAGCCCCAACCACTACACTCA 492  
Db 421 CACCCAGGCCCCAACCACTCTAGGCCCCAACCAACGAGCCCCAACCACTACACTCA 480  
QY 493 GGGCCCTACTACTACCACTAGGCCCCAACCACTCTAGGCCCCAACCACTACACTCA 552  
Db 481 GGGCCCTACTACTACCACTAGGCCCCAACCACTCTAGGCCCCAACCACTACACTCA 540  
QY 553 CCAGGCCCCAACCACTCTAGGCCCCAACCACTCTAGGCCCCAACCACTACACTCA 612  
Db 541 CCAGGCCCCAACCACTCTAGGCCCCAACCACTCTAGGCCCCAACCACTACACTCA 600  
QY 613 CCGCACTACATCACTCCAGGCTCACTACCTCCGCGCACTACCTCCGCGCACTAC 672  
Db 601 CCGCACTACATCACTCCAGGCTCACTACCTCCGCGCACTACCTCCGCGCACTAC 660  
QY 673 CCGGCGCGCACTACCTCCGCGCACTACCTCCGCGCACTACCTCCGCGCACTAC 732  
Db 661 CCGGCGCGCACTACCTCCGCGCACTACCTCCGCGCACTACCTCCGCGCACTAC 720  
QY 733 ACCCACTTACGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 792  
Db 721 ACCCACTTACGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAG 780  
QY 793 TGACTTGCAGATCACTTGTGATTCGCCAGCAAGTACTGCAACCTCTTCAACAGTG 852  
Db 781 TGACTTGCAGATCACTTGTGATTCGCCAGCAAGTACTGCAACCTCTTCAACAGTG 840  
QY 853 CTCCAACGCTTACACTTGCAGAGAGAGTGCCTGAGGAGCTTACTTCAACCTTACGT 912  
Db 841 CTCCAACGCTTACACTTGCAGAGAGAGTGCCTGAGGAGCTTACTTCAACCTTACGT 900  
QY 913 CCAGGCTGAGCTTCTCTCTAAGCTTGAAGTGAGAGGAGAAATCAGGCCCCGAGCCCC 972  
Db 901 CCAGGCTGAGCTTCTCTCTAAGCTTGAAGTGAGAGGAGAAATCAGGCCCCGAGCCCC 960  
QY 973 AGTCACGAAGAGGACGAAGAGCAATGACATGGAAGTGCAGACCTCTCTGACAATGATG 1032  
Db 961 AGTCACGAAGAGGACGAAGAGCAATGACATGGAAGTGCAGACCTCTCTGACAATGATG 1020  
QY 1033 CCGAGCTAATCTTGAATTCAGCTGCTTGGCCCCAGGAAACCGTTGCGACAATTTA 1092  
Db 1021 CCGAGCTAATCTTGAATTCAGCTGCTTGGCCCCAGGAAACCGTTGCGACAATTTA 1080  
QY 1093 CCAGTGGCTCCAGGTAATCTGTAAGAGAGGCTTGTGAGAGCGGAGCCCACTTCACTTT 1152  
Db 1081 CCAGTGGCTCCAGGTAATCTGTAAGAGAGGCTTGTGAGAGCGGAGCCCACTTCACTTT 1140  
QY 1153 TGAATCTCAGCAATGTGACCACTGAGCTGCTTGGCTGCACTCCCGCGCGAGAG 1212  
Db 1141 TGAATCTCAGCAATGTGACCACTGAGCTGCTTGGCTGCACTCCCGCGCGAGAG 1200  
QY 1213 CGAAGAACTTGAAGTGCAGAGAGTGCCTGACCGGCTGCTACTGCCCAAGGAACCAT 1272  
|||||

Db 1201 CGAAGAACTTGAAGTGCAGAGAGATGCTCAACCGGCTGTACTGCCCAAGGAACCAT 1260  
QY 1273 TGAATGGAGGCCCTTCCCAAGGCTGCCCTGCGCACTTCAGATGAGACACTCTCTCC 1332  
Db 1261 TGAATGGAGGCCCTTCCCAAGGCTGCCCTGCGCACTTCAGATGAGACACTCTCTCC 1320  
QY 1333 CCAGAGAGCGAGTGGGCGCAGTATCTACAGTGTGTCTATGACAGACTATGCAAGAC 1392  
Db 1321 CCAGAGAGCGAGTGGGCGCAGTATCTACAGTGTGTGTCTATGACAGACTATGCAAGAC 1380  
QY 1393 TTGCCCCGGAACCTGCACTTCACTGCTCCGACACAGTCTTGAAGTCTCTGTGACCGC 1452  
Db 1381 TTGCCCCGGAACCTGCACTTCACTGCTCCGACACAGTCTTGAAGTCTCTGTGACCGC 1440  
QY 1453 TGTGTCCAAAGTTTGTGAGTGGATTCAGACAACCGTGCAGATGCAATGCTGCCCCGAG 1512  
Db 1441 TGTGTCCAAAGTTTGTGAGTGGATTCAGACAACCGTGCAGATGCAATGCTGCCCCGAG 1500  
QY 1513 AGCTGTCCAAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGTGCAGACCTTC 1572  
Db 1501 AGCTGTCCAAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGTGCAGACCTTC 1560  
QY 1573 CACCGTGTCCCACTGCAAGCGCACCCGCACTGAGGCCCCAGTCCCACTCAACCGC 1632  
Db 1561 CACCGTGTCCCACTGCAAGCGCACCCGCACTGAGGCCCCAGTCCCACTCAACCGC 1620  
QY 1633 AATTCTCTCCGCGGCCCAAGGCTGCCCAAGGCTGCCCAAGTGCAGACCTTC 1692  
Db 1621 AATTCTCTCCGCGGCCCAAGGCTGCCCAAGGCTGCCCAAGTGCAGACCTTC 1680  
QY 1693 ATCCCAACCACTGTGCAAGTACCACTTGTGTGCCACGCGAGCCCTTACTACTG 1752  
Db 1681 ATCCCAACCACTGTGCAAGTACCACTTGTGTGCCACGCGAGCCCTTACTACTG 1740  
QY 1753 TGTGCTGGAATTCGCAATGCTGCAATGAGGCTTACCGGCTGCCCGCCAGCGCTCCCG 1812  
Db 1741 TGTGCTGGAATTCGCAATGCTGCAATGAGGCTTACCGGCTGCCCGCCAGCGCTCCCG 1800  
QY 1813 CACCGTGTCCCAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGTGCAGACCTTC 1872  
Db 1801 CACCGTGTCCCAAGGCTGCCCAAGGCTGCCCAAGGCTGCCCAAGTGCAGACCTTC 1860  
QY 1873 ATCAGCACTTACTGTGCCCCCACTACGCGAGCACTGCCCCCAACCAAGTCACTGT 1932  
Db 1861 ATCAGCACTTACTGTGCCCCCACTACGCGAGCACTGCCCCCAACCAAGTCACTGT 1920  
QY 1933 ACCAGCCACTGCT----- 1946  
Db 1921 ACCAGCCACTGCTGCCCTTACTACGCGAGCACTGCTGCCCAACCAAGTCACTGAC 1980  
QY 1946 -----GCCCTTACTACGCGAGCACTGCCCCCAACCAAGTCACTGAC 1995  
Db 1981 ACCAGCCAGGAGCCCTTACTACGCGAGCACTGCCCCCAACCAAGTCACTGAC 2040  
QY 1996 CACTGTGCCCCCACTGCACTGCCCCCACTGCCCCCAACCAAGTCACTGAC 2055  
Db 2041 CACTGTGCCCCCACTGCACTGCCCCCACTGCCCCCAACCAAGTCACTGAC 2100  
QY 2056 CCGGTAATCTACAAAGGAGGAGCACTACCACTGGAAGTGAAGTGAAGTGAAGTGA 2115  
Db 2101 CCGGTAATCTACAAAGGAGGAGCACTACCACTGGAAGTGAAGTGAAGTGAAGTGA 2160  
QY 2116 TCTGCCCAAGGAGCCCTTACTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2175  
Db 2161 TCTGCCCAAGGAGCCCTTACTGGAAGTGAAGTGAAGTGAAGTGAAGTGAAGTGA 2220  
QY 2176 CCGTGAATGCAAACTGAGTGTGAGAGGCAACCAAGTCTGAGTGTGCTC 2235  
Db 2221 CCGTGAATGCAAACTGAGTGTGAGAGGCAACCAAGTCTGAGTGTGCTC 2280  
QY 2236 TGAAGTCTCAGTCAACCCCACTTAAAGCTGTGACTGCTTGAAGCTGCGTTG 2295  
Db 2281 TGAAGTCTCAGTCAACCCCACTTAAAGCTGTGACTGCTTGAAGCTGCGTTG 2340  
|||||





OY	1796	-GGCCCCACGGTGGCCCCACCGCGTGGCCCCACGAGCGCTACTAGTGGTTCAG	1853
Db	366	CCCCCTCTCTGCTCTCGTGGCCCTCTGCCCCCTCTGCTCCCTCTGCTCTGCTGGCCCTC	307
OY	1854	AAATCCCACTACTGTCATCATCAACCACTACTGTCGCCCACTACCGAGACCTGGCC	1913
Db	306	CTGGCCCTCTGCTCGTCTCGTGGCCCTCTGCTCTGCTCTGCTGCGCCCTCTGCGCC	247
OY	1914	CCAAACACCAAGTCACTGTAACCAACCACTGTCGCCCTACTACCGAGACCTGGCCCA	1973
Db	246	CTCTCTCTGCTCTGCGCCCTCTGCGCCCTCTCTCTGCTGCTGCTGCTGCTGCTG	187
OY	1974	ACACCAAGTCACTGTAACCAACCAAGTGGCCCCCACTGCACTGCTCCCTACCGTGCAC	2033
Db	186	CCCCCTCTGCCCCCTCTGCCCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCCCCCTC	127
OY	2034	ATGCACCCCAACACCACAGCTGCGCCCGGTAACTCAACACAGCCACACAGCTACCACTG	2093
Db	126	CTGGCCCTCTGCGCCCTCTGCGCCCTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCGC	67
OY	2094	AAGATGATGACATGACCCCCCTCTGCCCAAGACCCCAATCAACCTTGCTTGAAGAT	2153
Db	66	CTCTGCTCTGCGCCCTCTGCGCTCTGCTGCTGCTCTGCTCTACACGTTGGGGGAT	7
OY	2154	GCAAC 2158	
Db	6	CGATC 2	

```

RESULT 10
US-08-970-900-7/c
Sequence 7, Application US/08970900
GENERAL INFORMATION:
APPLICANT: Masucci, Maria G.
TITLE OF INVENTION: FUSION PROTEINS HAVING INCREASED HALF-LIVES
NUMBER OF SEQUENCES: 91
CORRESPONDENCE ADDRESS:
ADDRESSEE: Kathleen M. Williams, Banner & Witcoff, Ltd.
STREET: One Financial Center
City: Boston
STATE: Massachusetts
COUNTRY: USA
Zip: 02111
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/970, 900
FILING DATE: 14-NOV-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/030,986
FILING DATE: 15-NOV-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/048,945
FILING DATE: 25-JUN-1997
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 3255/59831
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 803 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-970-900-7

```

Query Match	6.2%;	Score 153.4;	DB 28;	Length 803;
Best Local Similarity	53.1%;	Pred. NO. 1.3e-25;		
Matches 353;	Conservative 0;	Mismatches 306;	Indels 6;	Gaps 1;

QY	1500	CTGTGTCCTCCGACAGCGTGCCTCCAAAGGGGAGCCCAAGAGGCTGCGCCCAACGGGTGCGCCCA	1559
Db	666	CTCTGTGCCCCCTCTGTGCTGTGCCCCCTGTGCCCCCTGTGCTGTGCCCCCTGTGCCCCCTC	607
QY	1560	CTGCGCAGCCCTCCACAGCTGTGTCACACGTGACGCAAGCCGCACTGTGACGGCCCACTGC	1619
Db	606	CTCTGTGCTCTGCCCCCTCTGCCCCCTCTGCTGTGCTGTGCCCCCTCTGCTGTGCCCCCTC	547
QY	1620	CACCTACACACGGCAATTCTCTATCTCGGGGCCCCACGCGTGGCCCCCAACGGAGCTCTACTA	1679
Db	546	CTGCGCCCTCTGTGCCCCCTGCTGTGCCCCCTGTGCCCCCTGTGCCCCCTGTGCCCCCTC	487
QY	1680	CTGTGTGCCCCCTGAAATCCCAACACATGTACAGTACAGTACACTTACTGTCTGTCCACAGGAG	1739
Db	486	CTCTGTGCCCCCTCTGTGCTCTGTGCCCCCTGTGCTGTGCCCCCTCTGTGCCCCCTC	427
QY	1740	CCCCCTACTGTGTGTGTCCTCCGAATATCCCAATACATGTACATCAGCGGCTAAGCGT----	1796
Db	426	CTCTGTGCTCTGTGCCCCCTGTGCTGTGCCCCCTGTGCCCCCTGTGCCCCCTGTGCTGTG	367
QY	1796	--GCCGCCACCGCTGTGCCCCCAGCGCTGCCCCCAGCGACGCCCTTACTGTGTGTCCAG	1853
Db	366	CCCCCTCTCTGTGCTGTGCCCCCTGTGCCCCCTGTGCCCCCTCTGTGCTGTGCCCCCTC	307
QY	1854	AAATCCCAACTACTGTGTACATCATACCACTACTGTGTGCCCCCACTACGGAGACACTGCC	1913
Db	306	CTGCGCCCTCTGTGCTGTGCCCCCTGTGCCCCCTGTGCCCCCTCTGTGCCCCCTC	247
QY	1914	CCAAACACCAAGTACATGTATACACCAACCACTGTGCCCCCTTACTACGGAGACTGTGCCCA	1973
Db	246	CTCTGTGCTGTGCTGTGCCCCCTGTGCCCCCTGTGCTGTGCTGTGCCCCCTCTGTGCTGTG	187
QY	1974	ACACCAAGTACTGTATACCAACCACTGTGTGCCCCCACTGACACTCCCCCTATACGTCGAC	2033
Db	186	CCCCCTGTGCCCCCTCTGTGCCCCCTGTGCCCCCTGTGCCCCCTCTGTGCCCCCTC	127
QY	2034	ATGCACCAACACACACAGTGTGCCCCGCTACTACAAACGACGACACAGTATACACACCTG	2093
Db	126	CTGCGCTCTCTGTGCCCCCTGTGCCCCCTGTGCCCCCTGTGCCCCCTCTGTGCTGTG	67
QY	2094	AAGATGATGACATGACCCCCCTCTGTGCCACGACCCCACTCAACCCCTGTGCTGAAGAT	2153
Db	66	CTCTGTGCTGTGCCCCCTGTGCCCCCTGTGCCCCCTGTGCCCCCTGTGCCCCCTC	7
QY	2154	GCACAC	2158
Db	6	CGATC	2

RESULT 11  
 US-08-733-369A-112/c  
 : Sequence 112, Application US/08733369A  
 : GENERAL INFORMATION:  
 : APPLICANT: MASUCCI, Maria G.  
 : TITLE OF INVENTION: GLYCINE-CONTAINING SEQUENCES CONFERRING  
 : TITLE OF INVENTION: INVISIBILITY TO THE IMMUNE SYSTEM.  
 : NUMBER OF SEQUENCE: 123  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Kathleen M. Williams, Banner & Witcoff, Ltd.  
 : STREET: One Financial Center  
 : CITY: Boston  
 : STATE: Massachusetts  
 : COUNTRY: USA  
 : ZIP: 02111  
 :  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Diskette, 3.50 inch, 1.44 Mb storage  
 : COMPUTER: IBM PC compatible

```
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Wordperfect 6.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/733,369A
FILING DATE: 17 October, 1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/522,995
FILING DATE: 01-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/529,190
FILING DATE: 15-SEP-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: SE 95013249
FILING DATE: 10-APR-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/00876
FILING DATE: 10-APR-1996
ATTORNEY/AGENT INFORMATION:
NAME: Williams, Kathleen M.
REGISTRATION NUMBER: 34,380
REFERENCE/DOCKET NUMBER: 95-1391-D
TELECOMMUNICATION INFORMATION:
TELEPHONE: 617-345-9100
TELEFAX: 617-345-9111
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 798 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: other nucleic acid
US-08-733-369A-112
```

Query Match 6.2%; Score 153.2; DB 18; Length 798;  
Best Local Similarity 53.2%; Pred. No. 1.4e-25;  
Matches 351; Conservative 0; Mismatches 303; Indels 6; Gaps 1;

```
QY 1500 CTCTGCCCCGAGAGCTGCTCAAGGCTGCCCCAAGCGCTGCCCCAAGCGCTGCCCCAA 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 CTCTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
QY 1560 CTGCGGACCCCTGACCGGTGTCCACCTGCAAGCGCACCGGACACTGCAAGCCCACTGC 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 601 CTCTGCTCTGCGCCCTGCTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 542
QY 1620 CACCTCAACGCGAATCTCACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACTA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 541 CTGCCCCCTCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 482
QY 1680 CTGCTGCCCCGAGATGCCCAAGCACTGTACAGTACCACTACTGCTGCTGCTGCTGCTG 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 CTCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 422
QY 1740 CCCCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 362
QY 1796 --GCGCGCAACGCGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 302
QY 1854 AATGCCAATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 CTGCGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 242
QY 1914 CCAACACACAGTACGTACACACACACACACACACACACACACACACACACACACACAC 1973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 182
QY 1974 ACAACACAGTACGTACACACACACACACACACACACACACACACACACACACACACAC 2033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 122
```

```
QY 2034 ATGACCCCAACACACAGTGGCCCCGCTACTACAAACAGCGACGAGTACACACACTG 2093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 CTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 62
QY 2094 AAGATGATGACATGACACCCCTCTCCCAACGACCCCATCAACCTTGCGTTGAAGAT 2153
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 61 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2
```

## RESULT 12

```
US-08-970-900-4/c
; Sequence 4, Application US/08970900
; GENERAL INFORMATION:
; APPLICANT: Masucci, Maria G.
; TITLE OF INVENTION: FUSION PROTEINS HAVING INCREASED HALF-LIVES.
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kathleen M. Williams, Banner & Wilcoff, Ltd.
; STREET: One Financial Center
; CITY: Boston
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02111
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.44 MB storage
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Wordperfect 6.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/970,900
; FILING DATE: 14-NOV-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/030,986
; FILING DATE: 15-NOV-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/048,945
; FILING DATE: 25-JUN-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Williams, Kathleen M.
; REGISTRATION NUMBER: 34,380
; REFERENCE/DOCKET NUMBER: 3255/59831
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-345-9100
; TELEFAX: 617-345-9111
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 798 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: other nucleic acid
US-08-970-900-4
```

Query Match 6.2%; Score 153.2; DB 28; Length 798;  
Best Local Similarity 53.2%; Pred. No. 1.4e-25;  
Matches 351; Conservative 0; Mismatches 303; Indels 6; Gaps 1;

```
QY 1500 CTGCTGCCCCGAGAGCTGCTCAAGGCTGCCCCAAGCGCTGCCCCAAGCGCTGCCCCAA 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 661 CTCTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 602
QY 1560 CTGCGGACCCCTGACCGGTGTCCACCTGCAAGCGCACCGGACACTGCAAGCCCACTGC 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 601 CTCTGCTCTGCGCCCTGCTGCGCCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 542
QY 1620 CACCTCAACGCGAATCTCACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACTA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 541 CTGCCCCCTCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 482
QY 1680 CTGCTGCCCCGAGATGCCCAAGCACTGTACAGTACCACTACTGCTGCTGCTGCTGCTG 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

[illegible]

```

RESULT 13
PCT-US99-06742-14
Sequence 14, Application PC/TUS9906742
GENERAL INFORMATION:
APPLICANT: Ruffner, Duane E.
APPLICANT: Pierce, Michael L.
APPLICANT: Chen, Zhidong
TITLE OR INVENTION: Directed Antisense Libraries
FILE REFERENCE: T6678.PCT
CURRENT APPLICATION NUMBER: PCT/US99/06742
CURRENT FILING DATE: 1999-03-28
EARLIER APPLICATION NUMBER: 60/079,792
EARLIER FILING DATE: 1998-03-28
NUMBER OF SEQ ID NOS: 34
SOFTWARE: word97
SEQ ID NO 14
LENGTH: 8705
TYPE: DNA
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: pshuttle
PCT-US99-06742-14

```

Query Match	6.2%;	Score 151.6;	DB 1;	Length 8705;
Best Local Similarity	53.2%;	Pred. No. 8.6e-25;		
Matches 347;	Conservative	0;	Mismatches 299;	Indels 6;
				Gaps 1;

[illegible]

Db	7553	ctctctgcccctctgtctctgtgcccctctgtctctgtgcccctctgtgcccctctgtgcccctc	7622
QY	1740	CCCCCTACTAGCTGTCGTGTCCTGAATCCCAATCACTGTACATCAAGCGCTACCGCT---	1796
Db	7623	ctctctgtctctgtgcccctctctgtctctgtgcccctctgtgcccctctgtctgtctgtc	7682
QY	1796	--GGCCCCAGCGCTGCCCCAGCGCTGCCCCAGCGAGCCCCCTACTAGCTGTGCCAG	1853
Db	7683	cccctctctctgtctctgtgcccctctgtgcccctctgtgcccctctctgtctgtcctc	7742
QY	1854	AAATCCCAACTACTGTCATCATCAACCACTACTGTCGCCCCCACTACCGCAGACCTGCC	1913
Db	7743	ctgcccctctctctgtctctgtgcccctctctctctgtctctgtcctctctgtgccc	7802
QY	1914	CCAAACACCAAGTCACTGTACTGTACCAACCCACTGTCGCCCCCTACTACCGCAGACCTGCCCA	1973
Db	7803	ctctctctgtctctgtgcccctctgtgcccctctctgtctctgtcctctctctgtctgtc	7862
QY	1974	ACACACACAGTCACTGTACTGTACCAACCCACTGTCGCCCCCACTGACGACTCCCCCTACCGTCAC	2033
Db	7863	cccctctctgtgcccctctgtgcccctctctctgtctgtgcccctctctgtctgtcctc	7922
QY	2034	ATGCACCCACACCAACGCTGCCCCGGGTAACTTACAACGCGCAGCAGCTACACACTG	2093
Db	7923	ctgtgcccctctgtgcccctctgtgcccctctctgtctctgtcctctctctgtctgtc	7982
QY	2094	AAAGTGAATGACATGAGACCCCTCTCTCCCAAGACCCCACTAACCTTGGCG	2145
Db	7983	ctctctgtctctgtgcccctctgtgcccctctgtctctgtctctgtctctgtctc	8034

```

1      RESULT 14
2      US-08-407-128-15/c
3      ; Sequence 15, Application US/08407128
4      ; GENERAL INFORMATION:
5      ; APPLICANT: Godowsky, Paul J. Lokker, Nathalie A. Mark, Melanie R.
6      ; TITLE OF INVENTION: HEPATOCTYTE GROWTH FACTOR PROTEINASE DOMAIN VARIANTS
7      ; NUMBER OF SEQUENCES: 21
8      ; CORRESPONDENCE ADDRESS:
9      ; ADDRESSEE: Genentech, Inc.
10     ; STREET: 460 Point San Bruno Blvd
11     ; CITY: South San Francisco
12     ; STATE: California
13     ; COUNTRY: USA
14     ; ZIP: 94060
15     ;
16     ; COMPUTER READABLE FORM:
17     ; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
18     ; COMPUTER: IBM PC compatible
19     ; OPERATING SYSTEM: PC-DOS/MS-DOS
20     ; SOFTWARE: patin (Genentech)
21     ; CURRENT APPLICATION DATA:
22     ; APPLICATION NUMBER: US/08/407,128
23     ; FILING DATE: 18-MAY-1992

```

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

US-08-407-128-15

Query Match 6.2%; Score 151.6; DB 13; Length 10596;  
Best Local Similarity 53.2%; Pred. No. 9.3e-25;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

```
OY 1500 CTGCTGCCCCGAGAGTGTCTCCAGAGGCTGCCCAAGGCTGCCCAAGGCTGCCCA 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2867 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1560 CTGCGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2807 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1620 CACCTACAAACCGCAATTCCTACTCCGCGCCCAACGCGCTGCCCAAGGCTGCTACTA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2747 CTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1680 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2687 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2628
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1740 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2627 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1796 --GCCCGCAGCGCTGCGCCCGCAGCGCTGCCCAACGCGCTGCCCAAGGCTGCTGCT 1853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2567 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2508
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1854 AATCCCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2507 CTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1914 CCAACACAGCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2447 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1974 ACACGACAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2387 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2034 ATGACACCAACAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2327 CTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2094 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2267 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

RESULT 15  
US-08-409-770-15/c  
Sequence 15, Application US/08409770

GENERAL INFORMATION:  
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.  
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCTYTE GROWTH FACTOR VARIANTS  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genentech, Inc.  
STREET: 460 Point San Bruno Blvd  
CITY: South San Francisco  
STATE: California  
COUNTRY: USA  
ZIP: 94080  
COMPUTER READABLE FORM:  
MEDIUM TYPE: 5.25 inch, 360 kb floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patin (Genentech)  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/409,770  
FILING DATE:  
CLASSIFICATION: 435

PRIOR APPLICATION DATA:  
APPLICATION NUMBER:  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 755.1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-409-770-15

Query Match 6.2%; Score 151.6; DB 13; Length 10596;  
Best Local Similarity 53.2%; Pred. No. 9.3e-25;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

```
OY 1500 CTGCTGCCCCGAGAGTGTCTCCAGAGGCTGCCCAAGGCTGCCCAAGGCTGCCCA 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2867 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1560 CTGCGCAGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2807 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2748
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1620 CACCTACAAACCGCAATTCCTACTCCGCGCCCAACGCGCTGCCCAAGGCTGCTACTA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2747 CTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2688
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1680 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2507 CTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2448
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1914 CCAACACAGCAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2447 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2388
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 1974 ACACGACAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2387 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2328
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2034 ATGACACCAACAGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2327 CTGCGCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2268
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 2094 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 2267 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2216
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: July 18, 1999, 02:08:42  
Job time: 3036 sec



GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 07:48:19 ; Search time 72.5 Seconds  
(without alignments)  
3104.358 Million cell updates/sec

Title: US-09-294-663-1

Perfect score: 2455

Sequence: 1 GTAACGTTAGTGAAGAAAGAA.....AAAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY\_NUC

Searched: 176461 seqs, 45838279 residues

Database: Issued\_Patents\_NA:\*

1: /cgn2\_6/ptodata/1/ina/5A.COMB.seq:\*  
2: /cgn2\_6/ptodata/1/ina/5B.COMB.seq:\*  
3: /cgn2\_6/ptodata/1/ina/5C.COMB.seq:\*  
4: /cgn2\_6/ptodata/1/ina/5D.COMB.seq:\*  
5: /cgn2\_6/ptodata/1/ina/PCT059.COMB.seq:\*  
6: /cgn2\_6/ptodata/1/ina/backfiles1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	151.6	6.2	10596	1	US-07-884-811-15	Sequence 15, Appl
2	151.6	6.2	10596	1	US-07-885-971-15	Sequence 15, Appl
3	151.6	6.2	10596	1	US-08-087-783A-15	Sequence 15, Appl
4	151.6	6.2	10596	1	US-08-194-088B-15	Sequence 15, Appl
5	151.6	6.2	10596	4	US-08-194-087-15	Sequence 15, Appl
6	151.6	6.2	10596	5	PCT-US93-04648-15	Sequence 15, Appl
7	113	4.6	2793	3	US-08-209-747-1	Sequence 1, Appl
8	113	4.6	2793	3	US-08-458-298-1	Sequence 1, Appl
9	108.6	4.4	32207	4	US-08-770-379-20	Sequence 1, Appl
10	100.6	4.1	9551	1	US-08-056-200-93	Sequence 1, Appl
11	99.4	4.0	6530	5	PCT-US93-03993-1	Sequence 93, Appl
12	93	3.8	1984	1	US-07-885-970A-25	Sequence 25, Appl
13	93	3.8	1985	1	US-08-298-687A-25	Sequence 25, Appl
14	93	3.8	1985	1	US-08-298-829-25	Sequence 25, Appl
15	91.4	3.7	913	1	US-08-217-327-3	Sequence 25, Appl
16	91.4	3.7	913	1	US-07-885-970A-3	Sequence 3, Appl
17	91.4	3.7	913	1	US-08-298-687A-3	Sequence 3, Appl
18	91.4	3.7	913	1	US-08-530-797-2	Sequence 2, Appl
19	91.4	3.7	913	1	US-08-298-829-3	Sequence 3, Appl
20	85.6	3.5	1505	1	US-07-815-246-1	Sequence 1, Appl
21	83.4	3.4	44377	4	US-08-804-227C-7	Sequence 7, Appl
22	83.4	3.4	13987	4	US-08-804-227C-13	Sequence 13, Appl
23	82.2	3.3	12001	3	US-08-458-568A-11	Sequence 11, Appl
24	81	3.3	7218	2	US-08-232-463-14	Sequence 14, Appl
25	77.6	3.2	2338	2	US-08-425-069-1	Sequence 1, Appl
26	77.2	3.1	10747	5	PCT-US93-03985-1	Sequence 1, Appl
27	77.2	3.1	24979	5	PCT-US93-03985-3	Sequence 3, Appl
28	76.4	3.1	2427	4	US-08-678-039A-39	Sequence 39, Appl
29	74.8	3.0	2242	4	US-08-641-627A-37	Sequence 37, Appl
30	71.8	2.9	43280	2	US-08-804-227C-1	Sequence 1, Appl
31	69.8	2.8	32207	4	US-08-770-379-20	Sequence 20, Appl
32	65.8	2.7	31571	2	US-08-323-443B-1	Sequence 18, Appl
33	65.4	2.7	3833	5	US-08-917-320-18	Sequence 18, Appl
34	65.4	2.7	3833	5	PCT-US95-04611A-18	Sequence 18, Appl
35	64.8	2.6	786	3	US-08-403-379A-2	Sequence 18, Appl
36	64.6	2.6	2093	3	US-08-287-001A-1	Sequence 1, Appl
37	64.6	2.6	2093	5	PCT-US95-09941-1	Sequence 1, Appl

C	38	63.4	2.6	234	3	US-08-469-802B-3	Sequence 3, Appl
	39	63.4	2.6	234	4	US-08-267-803B-3	Sequence 5, Appl
	40	63.2	2.6	2032	5	PCT-US95-07721-5	Sequence 5, Appl
	41	62.6	2.5	2753	4	US-07-952-853-5	Sequence 5, Appl
	42	61.8	2.5	10366	2	US-08-246-982A-5	Sequence 5, Appl
	43	61.8	2.5	10366	4	US-08-453-265-5	Sequence 5, Appl
	44	61.8	2.5	10346	2	US-08-457-273B-41	Sequence 41, Appl
	45	60.6	2.5	3624	1	US-07-951-715A-6	Sequence 6, Appl

## ALIGNMENTS

```
RESULT 1
US-07-884-811-15/c
Sequence 15, Application US/07884811
Patent No. 5316921
GENERAL INFORMATION:
APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.
TITLE OF INVENTION: SINGLE-CHAIN HEPATOCTYTE GROWTH FACTOR VARIANTS
NUMBER OF SEQUENCES: 21
CORRESPONDENCE ADDRESS:
ADDRESSEE: Genentech, Inc.
STREET: 460 Point San Bruno Blvd
CITY: South San Francisco
STATE: California
COUNTRY: USA
ZIP: 94080
COMPUTER READABLE FORM:
MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: patin (Genentech)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/884,811
FILING DATE: 19920518
CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Dreger, Ginger R.
REGISTRATION NUMBER: 33,055
REFERENCE/DOCKET NUMBER: 755.1
TELECOMMUNICATION INFORMATION:
TELEPHONE: 415/225-3216
TELEFAX: 415/952-9881
TELEX: 910/371-7168
INFORMATION FOR SEQ ID NO: 15:
SEQUENCE CHARACTERISTICS:
LENGTH: 10596 bases
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
US-07-884-811-15

Query Match 6.2%; Score 151.6; DB 1; Length 10596;
Best Local Similarity 53.2%; Pred. No. 3e-23;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;
```





CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/087,783A  
FILING DATE: 13-Jul-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/884811  
FILING DATE: 18-MAY-92  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 07/885971  
FILING DATE: 18-MAY-92  
ATTORNEY/AGENT INFORMATION:  
NAME: Marschang, Diane L.  
REGISTRATION NUMBER: 35,600  
REFERENCE/DOCKET NUMBER: P0755779P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-5416  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 base pairs  
TYPE: Nucleic Acid  
STRANDEDNESS: Single  
TOPOLOGY: Linear  
US-08-087-783A-15

Query Match 6.2%; Score 151.6; DB 1; Length 10596;

Best Local Similarity 53.2%; Pred. No. 3e-23; Mismatches 299; Indels 6; Gaps 1;

Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

OY 1500 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1559  
DB 2867 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2808  
OY 1560 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1619  
DB 2807 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2748  
OY 1620 CACCTACACGCGCAATCTCTACTCGGCGCGCCCAACGGCTGCCCAAGGCTGCCCA 1679  
DB 2747 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2688  
OY 1680 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1739  
DB 2687 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2628  
OY 1740 CCGCTACTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1796  
DB 2627 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2568  
OY 1796 --GCCCGCAGCGCTGCCCGCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1853  
DB 2567 CCGCTACTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2508  
OY 1854 AATCCAGCTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1913  
DB 2507 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2448  
OY 1914 CCAACAGCAGTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1973  
DB 2447 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2388  
OY 1974 ACACAGCAGTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2033  
DB 2387 CCGCTACTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2328  
OY 2034 ATGCAACCAACAGCAGTGGCGGCTACTACAACGAGCAGCAGTACCAACAGT 2093  
DB 2327 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2268  
OY 2094 AAGATGATGACATGACAGCGGCTGCTCCAGCGGCGCCCAACGGCTGCCCA 2145

DB 2267 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2216

#### RESULT 4

US-08-194-088B-15/C

; Sequence 15, Application US/08194088B

; Patent No. 5580963

; GENERAL INFORMATION:

; APPLICANT: Godowski, Paul J. Lokker, Nathalie A. Mark, Melanie R.

; TITLE OF INVENTION: SINGLE-CHAIN HEPATOCYTE GROWTH FACTOR VARIANTS

; NUMBER OF SEQUENCES: 21

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Genentech, Inc.

; STREET: 460 Point San Bruno Blvd

; CITY: South San Francisco

; STATE: California

; COUNTRY: USA

; ZIP: 94080

; COMPUTER READABLE FORM:

; MEDIUM TYPE: 5.25 inch, 360 KB floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: patin (Genentech)

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/194,088B

; FILING DATE: 09-FEB-1994

; CLASSIFICATION: 530

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/884811

; FILING DATE: 18-MAY-1992

; ATTORNEY/AGENT INFORMATION:

; NAME: Gallegos, R. Thomas

; REGISTRATION NUMBER: 32,692

; REFERENCE/DOCKET NUMBER: 755D1

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 415/225-2614

; TELEFAX: 415/952-9881

; TELEX: 910/371-7168

; INFORMATION FOR SEQ ID NO: 15:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 10596 bases

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; US-08-194-088B-15

Query Match 6.2%; Score 151.6; DB 1; Length 10596;

Best Local Similarity 53.2%; Pred. No. 3e-23; Mismatches 299; Indels 6; Gaps 1;

Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

OY 1500 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1559  
DB 2867 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2808  
OY 1560 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1619  
DB 2807 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2748  
OY 1620 CACCTACACGCGCAATCTCTACTCGGCGCGCCCAACGGCTGCCCAAGGCTGCCCA 1679  
DB 2747 CTGCGCGCAGCTCCAGCGGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2688  
OY 1680 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1739  
DB 2687 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2628  
OY 1740 CCGCTACTACTGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1796  
DB 2627 CTGCTGCCCGCAGCTGCTCCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 2568  
OY 1796 --GCCCGCAGCGCTGCCCGCAGCGGCGCCCAACGGCTGCCCAAGGCTGCCCA 1853



NAME: Dreger, Ginger R.  
REGISTRATION NUMBER: 33,055  
REFERENCE/DOCKET NUMBER: 755,779P1  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 415/225-3216  
TELEFAX: 415/952-9881  
TELEX: 910/371-7168  
INFORMATION FOR SEQ ID NO: 15:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10596 bases  
TYPE: NUCLEIC ACID  
STRANDEDNESS: single  
TOPOLOGY: linear  
PCT-US93-04548-15

Query Match 6.2% Score 151.6; DB 5; Length 10596;  
Best Local Similarity 53.2%; Pred. No. 3e-23;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

QY 1500 CTGCTGCCCCGACAGCTGCTCCAGAGGCTGCCCAAGGCTGCCCAAGGCTGCCCA 1559  
DB 2867 CTGCTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808  
QY 1560 CTGCGGACCTTCACCGTGTGCTCCACTGCAAGCCGACCCGCACTGCAGGCCAGTCC 1619  
DB 2807 CTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2748  
QY 1620 CAGCTACAGCGCAATTCCTAGCTGCGGCGCCGACGCTGCCGACGCTGCTGCTGCTA 1679  
DB 2747 CTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2688  
QY 1680 CTGCTGCCCCCTGCAATCCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1739  
DB 2687 CTGCTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2628  
QY 1740 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796  
DB 2627 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2568  
QY 1796 --GCCGCCACCGCTGCCCGCCGCTGCCCGCCGCTGCCCGCCGCTGCCCGCTGCC 1853  
DB 2567 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2508  
QY 1854 AATGCCAATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913  
DB 2507 CTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2448  
QY 1914 CCAAGACAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1973  
DB 2447 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2388  
QY 1974 ACACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2033  
DB 2387 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2328  
QY 2034 ATGACGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093  
DB 2327 CTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2268  
QY 2094 AAGATGATGACATGACGCGCCCTGCTGCCAGACGACCCGATCAACCTGCTGCT 2145  
DB 2267 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2216

RESULT 7  
US-08-209-747-1/c  
Sequence 1, Application US/08209747  
Patent No. 5733771  
GENERAL INFORMATION:  
APPLICANT: Lewis, Randolph V.  
APPLICANT: Colgin, Mark  
TITLE OF INVENTION: CDNAS Encoding Minor Ampullate Spider

TITLE OF INVENTION: Silk Proteins  
NUMBER OF SEQUENCES: 56  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Birch, Stewart, Kolasch & Birch  
STREET: P.O. Box 747  
CITY: Falls Church  
STATE: Virginia  
COUNTRY: USA  
ZIP: 22040-3487  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/209,747  
FILING DATE: 14-MAR-1994  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Murphy Jr., Gerald M.  
REGISTRATION NUMBER: 28,977  
REFERENCE/DOCKET NUMBER: 1447-104P  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-205-8050  
FAX: 703-205-8050  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2793 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ORIGINAL SOURCE:  
ORGANISM: Nephila clavipes  
TISSUE TYPE: minor ampullate gland  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 183..2675  
OTHER INFORMATION: /product="N. clavipes minor  
OTHER INFORMATION: ampullate silk protein"  
US-08-209-747-1

Query Match 4.6% Score 113; DB 3; Length 2793;  
Best Local Similarity 50.9%; Pred. No. 2.3e-15;  
Matches 269; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

QY 1494 CATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1553  
DB 2120 CAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2061  
QY 1554 CCGCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1613  
DB 2060 CTGCGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2001  
QY 1614 CAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1673  
DB 2000 CTGGAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1941  
QY 1674 CTACTAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1733  
DB 1940 CAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1881  
QY 1734 CCGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1793  
DB 1880 CAGCTGCGGACGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1821  
QY 1794 CTGCCCCGACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1853  
DB 1820 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1761  
QY 1854 AATGCCAATCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913



LENGTH: 32207 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-770-379-20

Query Match 4.4%; Score 108.6; DB 4; Length 32207;  
Best Local Similarity 48.5%; Pred. No. 3.7e-14;  
Matches 300; Conservative 0; Mismatches 319; Indels 0; Gaps 0;

QY 1503 CTGCCCCGACAGCTGCTCCACAGGCTGCCCAACGGCTGCCCCCAAGGCTGCCCACTG 1562  
DB 19348 CTCCCTTACTC 19407  
QY 1563 CCGACCCCTCCACGGTGTCCACCTGCACAGCGCCACCGCACTGACCCCAAGTCCAC 1622  
DB 19408 TAATTCCTGCTC 19467  
QY 1623 CTACACCGCAATTCCTC 1682  
DB 19468 CTCTGCTC 19527  
QY 1683 CTGCCCCGACAGCTGCTCCACAGTACAGTACAGTACAGTACAGTACAGTACAGTACAG 1742  
DB 19528 CTGCT 19587  
QY 1743 CTACTACTGCTGTCT 1802  
DB 19588 CTCTGCTC 19647  
QY 1803 CCGCTGCCCCACCGCTGCTCCCAACGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTC 1862  
DB 19648 TAATTCCTGCTC 19707  
QY 1863 CTACTGCTACATCACCACCT 1922  
DB 19708 CTCTGCTG 19767  
QY 1923 CAGTACTGCTACACCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1982  
DB 19768 CTGCTGCTG 19827  
QY 1983 TCAGTACTGCTACACCGCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 2042  
DB 19828 CTGCTGCTC 19887  
QY 2043 ACACCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2102  
DB 19888 CTGCTGCTC 19947  
QY 2103 ACATCGACCCCT 2121  
DB 19948 ATCTGCTC 19966

RESULT 10  
US-08-056-200-93/C

Sequence 93, Application US/08056200  
Patent No. 5616500  
GENERAL INFORMATION:  
APPLICANT: Steinert, Peter M.  
APPLICANT: Lee, Seung-Chul  
APPLICANT: Kim, In-Gyu  
APPLICANT: Chung, Soo-Il  
APPLICANT: Park, Sang-Chul  
TITLE OF INVENTION: Trichohyalin and Transglutaminase-3 and  
NUMBER OF INVENTION: Methods of Using Same  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Knobbe, Martens, Olson & Bear  
STREET: 620 Newport Center Drive, Sixteenth Floor

CITY: Newport Beach  
STATE: CA  
COUNTRY: U.S.A.  
ZIP: 92660  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
OPERATING SYSTEM: IBM PC compatible  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/056,200  
FILING DATE: 30-APR-1993  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Fedrick, Michael F.  
REGISTRATION NUMBER: 36,799  
REFERENCE/DOCKET NUMBER: NIH054.001A  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (714) 760-0404  
TELEFAX: (714) 760-9502  
INFORMATION FOR SEO ID NO: 93:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 9551 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: cDNA  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 1507..1644  
FEATURE:  
NAME/KEY: Intron  
LOCATION: 1645..2511  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 2512..8070  
US-08-056-200-93

Query Match 4.1%; Score 100.6; DB 1; Length 9551;  
Best Local Similarity 48.7%; Pred. No. 1.2e-12;  
Matches 274; Conservative 0; Mismatches 289; Indels 0; Gaps 0;

QY 1494 CATGACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1553  
DB 3771 CAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 3712  
QY 1554 CCGCAACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1613  
DB 3711 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 3652  
QY 1614 CAGTCCACCTACACCGCAATTCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1673  
DB 3651 CAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3592  
QY 1674 CTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1733  
DB 3591 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 3532  
QY 1734 CCGGAGCCCTACTACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1793  
DB 3531 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 3472  
QY 1794 CTGCCCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1853  
DB 3471 GCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3412  
QY 1854 AAATCCACCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913  
DB 3411 CCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTC 3352



LENGTH: 1984 base pairs  
TYPE: NUCLEIC ACID  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Gossypium barbadense  
STRAIN: Sea Island  
IMMEDIATE SOURCE:  
LIBRARY: EMBL SI  
CLONE: SIH6  
US-07-885-970A-25

Query Match 3.8% Score 93; DB 1; Length 1984;  
Best Local Similarity 53.1%; Pred. No. 3.2e-11;  
Matches 198; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1593 CGCCACCCGCAATGCGAGCCCGCCAGCTACACGCAATTCCTACTCGGCCCCCA 1652  
DB 424 CGCCACCCGCAACAGCGGCTTCTACTCTCTCGAGCTCAAGCAGCGGCTACACCA 483  
QY 1653 CGGCTGCCCCCAGCGAGCTCTACTACTGTGCCCCGTAATCCCAACCACTGTACAG 1712  
DB 484 CGGCACTCCGCGACGCTTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 543  
QY 1713 TACCACTACTGTGCTCCAGCGCCGCTACTACTGTGCTCTGTAATCCCAATTA 1772  
DB 544 CTCTTCACCCCGAGTTTAACTCTCTCCAGCGAGTTCTCTCTCTCTCTCTCTCTCT 603  
QY 1773 CTGTCAATCAGCGCTACCGCTGCCCCGCGCCGCTGCCCCGCGAG 1832  
DB 604 CTGCTTCT 663  
QY 1833 CCCCTACTACTGTGTCCAGAAATCCCACTACTGTACATCAGCACTACTGTGCCC 1892  
DB 664 CAGCTTCT 723  
QY 1893 CCAGTACCGCAGACCTGCCCCCAACACCACTACTGTACCACTACTGTGCCCCCTA 1952  
DB 724 CTGTACCCCGCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 783  
QY 1953 CTACCGCAGCACC 1965  
DB 784 CAGTACCAACACC 796

RESULT 13  
US-08-298-687A-25

; Sequence 25, Application US/08298687A  
; Patent No. 5521078

; GENERAL INFORMATION:  
; APPLICANT: John, Maliyaka E.  
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON  
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER  
; NUMBER OF SEQUENCES: 33  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Nicholas J. Seay, Quarles & Brady  
; STREET: P.O. Box 2113, First Wisconsin Plaza  
; CITY: Madison  
; STATE: Wisconsin  
; COUNTRY: USA  
; ZIP: 53701  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Microsoft Word  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/298,687A  
; FILING DATE:

CLASSIFICATION: 800  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/617,239  
FILING DATE: 21-NOV-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/253,243  
FILING DATE: 04-OCT-1988  
ATTORNEY/AGENT INFORMATION:  
NAME: Seay, Nicholas J.  
REGISTRATION NUMBER: 27,386  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (608) 283-2478  
TELEFAX: (608) 251-5139  
INFORMATION FOR SEQ ID NO: 25:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1985 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Gossypium barbadense  
STRAIN: Sea Island  
IMMEDIATE SOURCE:  
LIBRARY: EMBL-SI  
CLONE: SIH6  
US-08-298-687A-25

Query Match 3.8% Score 93; DB 1; Length 1985;  
Best Local Similarity 53.1%; Pred. No. 3.2e-11;  
Matches 198; Conservative 0; Mismatches 175; Indels 0; Gaps 0;

QY 1593 CGCCACCCGCAATGCGAGCCCGCCAGCTACACGCAATTCCTACTCGGCCCCCA 1652  
DB 424 CGCCACCCGCAACAGCGGCTTCTACTCTCTCGAGCTCAAGCAGCGGCTACACCA 483  
QY 1653 CGGCTGCCCCCAGCGAGCTCTACTACTGTGCCCCGTAATCCCAACCACTGTACAG 1712  
DB 484 CGGCACTCCGCGACGCTTCTACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 543  
QY 1713 TACCACTACTGTGCTCCAGCGCCGCTACTACTGTGCTCTGTAATCCCAATTA 1772  
DB 544 CTCTTCACCCCGAGTTTAACTCTCTCCAGCGAGTTCTCTCTCTCTCTCTCTCTCT 603  
QY 1773 CTGTCAATCAGCGCTACCGCTGCCCCGCGCCGCTGCCCCGCGAG 1832  
DB 604 CTGCTTCT 663  
QY 1833 CCCCTACTACTGTGTCCAGAAATCCCACTACTGTACATCAGCACTACTGTGCCC 1892  
DB 664 CAGCTTCT 723  
QY 1893 CCAGTACCGCAGACCTGCCCCCAACACCACTACTGTACCACTACTGTGCCCCCTA 1952  
DB 724 CTGTACCCCGCAGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 783  
QY 1953 CTACCGCAGCACC 1965  
DB 784 CAGTACCAACACC 796

RESULT 14

US-08-298-829-25  
; Sequence 25, Application US/08298829

; Patent No. 5620882  
; GENERAL INFORMATION:  
; APPLICANT: John, Maliyaka E.  
; TITLE OF INVENTION: GENETICALLY ENGINEERING COTTON  
; TITLE OF INVENTION: PLANTS FOR ALTERED FIBER  
; NUMBER OF SEQUENCES: 33





```
Db 175 CGACACCCACACACCGGCTTCTACTCTCTCCGACGACTCAAGCACCGCCTACACCAA 234
OY 1653 CCGCTGCCCCCAACCGGAGCTCTACTACTGCTGCCCCGATATCCCAACCACTGTACAG 1712
Db 235 CCGCCACTCCGGCACCGGTTTCTACTCTCTCCACTTCATCACCAGCCCCCAGTGACAG 294
OY 1713 TACCACTACTGCTGCTCCACCGACGCCCTACTACTGCTGCTCCCTGAATCCCAATCA 1772
Db 295 CTTCTCCACCCCACTTTCAACTCCTCCACCCAGTTCTCTCCTCCTGCACTCCACAC 354
OY 1773 CTGTACATCAAGCGCTTACGCTGCCCCCAACCGCTGCCCCCAACCGCTGCCCCCAAC 1832
Db 355 CTGCTTCTCTCTCTCTGCACTCCGACCTTCCTCCTGCACTCCGCTGCTGCTGCTG 414
OY 1833 CCCCTACTACTGCTGCTGCTCCAGAAATCCCAACTACTGCTACATCACCACCTACTGCTG 1892
Db 415 CAGCTTCTCCACCTCCCGCACTCCACCACTGCAACCCACCGCCAGCAACTCTCTCTC 474
OY 1893 CCACCTACCGCAGCAGCTGCCCCCAACCAACCAAGTCACGTACACCACTGCTGCCCC 1952
Db 475 CTGCTACCCCAACCAAGCAGTTCATGCTTCTCTCTCCAGCAAGTCACGATATCTCTC 534
OY 1953 CTACCGCAGCACC 1965
Db 535 CAGTACAAACACC 547
```

Search completed: July 18, 1999, 07:53:30  
Job time: 311 sec

**THIS PAGE BLANK (uspto)**

PKS 741 insert com  
Intestinal mucin c

Sequence 799 BP; 201 A; 106 C; 479 G; 13 T;

Best Local Similarity 53.2%; Pred.No. 1./e-21;  
Matches 352; Conservative 0; Mismatches 304; Indels 6; Gaps 1

1500 CTGCTGCCCGACAGCTGCTCCACGCGCTGCCCCACGCGTGCCCA 1559

Db 662 CTCTGCCCCCTCCTGCTCCTGCCCCCTCCTGCTCCTGCTCCTGCCCCCTC 603

1560 CTGCCGACCCCTCCACCGTGGTCCCACTGCAACGCCACCCGCACTGCAGCCCCAGTCC 1619

D<sub>b</sub>      602 CTCCTGCTCCTGCCCCCTCCTGCCTCCTGCCCCCTCCTCGTC    543

1620 CACCTACACCGCAATTCCTACTCCGGCCCCCACCAGCTGCCCTACTA 1675

Db 542 CTGCCCCCTCCTGCCCCCTGCTCCTGCCCCCTCCTGCTCCTGCCCCCTCCTG 483

QY 1680 CTGCTGCCCTGATCCCCAACCACCTGTACACGTACCACTACTGCTGCCACCGCAG 1739

Db 482 CTCCTGCCCTCCTGCTCCTGCCCTCCTGCTCCTGCCCTCCTGCCCCC 423

1740 CCCCTACTGCTGTCCCTGAATCCCAATCACTGTACATCAGGCCCTACCGCT--- 1796

Chr	Start (kb)	End (kb)	Feature	Annotation
Db	422	422	CTCTGCTGCTGCGCCCTCTGCTGCTGCGCCCTCTGCGCCCTCTGCGCCCTCTGCTGCTG	363
Oy	1796	--GCCCCCAACCGCTGCCCCCAACCGCTGCCCCCAACCGCAACCCCTACTACTGCTGTCCAG	18533	
Db	362	CCCCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	303	
Oy	1854	AAATCCCAACTACTGCTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	191313	
Db	302	CTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	243	
Oy	1914	CCAAACACCAAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	19737	
Db	242	CTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	183	
Oy	1974	ACACCAACAGTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT	20333	
Db	182	CCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	123	
Oy	2034	ATGCAACCAACCTG	20933	
Db	122	CTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	63	
Oy	2094	AAGATGATGACATGACATGACATGACATGACATGACATGACATGACATGACATGACATGACAT	2153	
Db	62	CT	3	
Oy	2154	GC 2155		
Db	2	CC 1		
RESULT	2			
O51731/C				
ID	O51731	standard; DNA; 10596 BP.		
AC	O51731:			
DT	31-MAY-1994	(first entry)		
DE	Plasmid pCISBON for subcloning huGF variants.			
KW	Hepatocyte Growth Factor; HGF; variant; mutin; in vitro mutagenesis;			
KW	proteolysis resistant; liver; malignancy; CMV-driven;			
OS	Cytomegalovirus; episomal expression plasmid; ss.			
OS	Synthetic.			
PH	Key			
FT	enhancer	Location/Qualifiers		
FT		1..611		
FT	/tag= a			
FT	/note= "CMV enhancer/promoter"			
FT	758..775			
FT	/tag= b			
FT	/label= SP6-promoter			
FT	845..849			
FT	/tag= c			
FT	/note= "SP6 RNA start"			
FT	902..966			
FT	/tag= d			
FT	/function= cloning-linker			
FT	967..1107			
FT	/tag= e			
FT	/note= "SV40 poly A"			
FT	1108..1531			
FT	/tag= f			
FT	/function= SV40-origln			
FT	1580..4189			
FT	/tag= g			
FT	/label= EBNA-1			
FT	4190..6374			
FT	/tag= h			
FT	/function= orip			
FT	4295..4887			
FT	/tag= i			
FT	/note= "family of repeats"			
FT	5866..5978			
FT	/tag= j			
FT	/note= "dyad region"			

FT	terminator	6375..6457
FT		/+tag- k
FT		/label= HSV_TK-terminator_3'-end
FT	cds	6975..7975
FT		/+tag- 1
FT		/phenotype= neomycin_resistance
FT		/note= "n5 neomycin phosphotransferase gene"
FT	promoter	7975..8112
FT		/+tag- m
FT		/label= TK-Promoter
FT		8114..8594
FT		/+tag- n
FT		/function= M13_ori
FT		8595..10414
FT	misc_rna	
FT		/+tag- o
FT		/label= delta_2a
PN		
PN	MO9323541-A.	
PD	25-NOV-1993.	
PF	17-MAY-1993.	U04648.
PR	18-MAY-1992.	US-8848411.
PR	18-MAY-1992.	US-885971.
PA	(GETH.) GENENTECH INC.	
PI	Godowski PJ. LOKKER.NA.	Mark MR;
DR	MPR. 93-386573/48.	
PT	Hepatocyte growth factor variants -	are resistant to proteolytic
PT	Cleavage into its two-chain form,	used to treat malignancies
PT	associated with HGF receptor	
PS	Example 1; Fig 6; 87pp; English.	
CC	Plasmid pcIEBON (a pRK5 derivative)	is an episomal CMV driven
CC	expression plasmid. bHGF variants	with enhanced receptor binding
CC	activity were produced by site-directed	mutagenesis. Stable
CC	populations of preferred HGF variants	were obtained by transfecting
CC	human embryonic kidney 293 cells and	then these were subcloned in
CC	pcIEBON. See R52940-R52949, for	examples of pref. HGF variants.
Q	Sequence 10596 BP; 2625 A; 2571 C;	3024 G; 2376 T;

Query Match	6-28;	Score 151.6;	DB 1;	Length 10596;
Best Local Similarity	53.28;	Pred. No. 7e-21;		
Matches 347;	Conservative	0;	Mismatches 299;	Indels 6; Gaps 1

[illegible]

OY	1974	ACACCAGATCACTGTAACCCAGCAGTGGCCCCCACTGAGTGTCCCCTACCGTGAC	2033
Db	2387	CCCCCTCTGCCCTCTGTCGCCTCTCTCTGCTCTGTCGCCTCTCTCTCTGCTCTG	2228
OY	2034	ATGCACCCAAACAACAGCTGCGCCCGGTACTAACAACAGCGCACAGTACCAACTG	2093
Db	2327	CTGCCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2268
OY	2094	AAGATGAATGACATCGAACCCCTCTCTCCCAAGACCCCATCAACCTTGCT	2145
Db	2267	CTCTGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2216
 RESULT 3			
ID	T40348/c	T40348 standard; DNA; 10596 BP.	
AC	T40348;		
DT	09-DEC-1996	(first entry)	
DE	Plasmid pcISEBON for expression of hepatocyte growth factor.		
KW	Human; hepatocyte growth factor; HGF; hHGF; serum; proteolytic cleavage;		
KM	pro-hormone; beta subunit; alpha subunit; kringe domain; prothrombin;		
KW	plasmaHomo; catalytic domain; serine protease; HGF variant;		
KM	HGF receptor; malignancy; chronic HGF receptor activation; ss.		
OS	Synthetic.		
PN	US5547856-A.		
PD	20-AUG-1996.		
PF	18-MAY-1992; 884811.		
PR	18-MAY-1992; US-885971.		
PR	18-MAY-1992; US-884811.		
PA	13-JUL-1993; US-087783.		
PR	(GENE) GENENTECH INC.		
PI	Godowski PJ, Lorker NA, Mark MR:		
DR	WPI; 96-352634/39.		
PT	New hepatocyte growth factor variants - are resistant to in vivo		
PS	proteolytic cleavage into a 2-chain form, useful as HGF antagonists		
CC	Example 1; Fig 6; 39pp; English.		
CC	This sequence represents the episomal CMV driven expression plasmid		
CC	pcISEBON which was used in the expression of variant human hepatocyte		
CC	growth factor (HGF). HGF is isolated from human serum and is a		
CC	disulphide linked heterodimer derived by proteolytic cleavage of the		
CC	pro-hormone between residues 494 and 495. This generates a molecule		
CC	composed of an alpha subunit of 440 amino acids (mol. wt. 69 kD) and		
CC	a beta subunit of 224 amino acids (mol. wt. 34 kD). The alpha and beta		
CC	subunits are encoded by a single open reading frame. The alpha subunit		
CC	contains four kringe domains based on their homology to kringe-like		
CC	domains in other proteins, e.g. prothrombin, plasminogen. The beta		
CC	subunit shows high homology to the catalytic domain of serine proteases.		
CC	However two of the three residues which form the catalytic triad of		
CC	serine proteases are not conserved in HGF. Therefore, the precise		
CC	function of the beta chain remains unknown. The invention includes HGF		
CC	variants which retain HGF receptor binding activity without having the		
CC	biological activity of wild-type HGF. They can be used for the treatment		
CC	of pathological conditions associated with the activation of a HGF		
CC	receptor such as pcISEBON plasmid comprising with chronic HGF receptor		
CC	activation. The pcISEBON plasmid comprising the HGF coding sequence		
CC	may be used for manipulation of the HGF coding sequence and expression		
CC	of the variant HGF's of the invention.		
SO	Sequence 10596 BP; 2627 A; 2571 C; 3023 G; 2375 T;		
 Query Match 6.2%; Score 151.6; DB 1; Length 10596; Best Local Similarity 53.2%; Pred. No./e-21; Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1			
OY	1500	CTGCTGCCCGGACACTGCTCCACGAGCGGCCCAACGCGCTGCCCCCAAGCGGTGCCCAA	1559
Db	2867	CTCTGCTGCCCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2808
OY	1560	CTGCGGCAACCTTCACCGTGTCCCACTGCAACGCGCACCGGCACTGCAAGCCCCAGTCC	1619
Db	2807	CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	2748
OY	1620	CACCTAACAACCGCAATTCTACTCGGCGGCCCAACGCGCTGCCCCCAACGCAAGCTCTACTA	1679

[illegible]

PA (CHIR) CHIRON CORP.  
 PI Cohen F, Dubois-Stringfellow N, Dwarki V, Innis MA,  
 PI Murphy JE, Tetsuo U, Zukermann R;  
 DR WPI: 98-159296/14.  
 PT Polyclonetic agents based on alpha-amino acids, able to complex  
 PT with nucleic acid - to facilitate its entry into cell, condense it  
 PT and protect it against serum degradation, particularly for use in  
 PT gene therapy.  
 PS Disclosure: Page 77-80; 100pp; English.  
 CC This polynucleotide comprises the DNA sequence of vector plasmid  
 CC PCMVKntR-EpI, which contains an Epstein-Barr virus (EBV) origin  
 CC of replication from plasmid pCEP4, a coding region for EBV nuclear  
 CC antigen A from pCEP4, a pair of inverted terminal repeats from  
 CC adeno-associated virus, a cytomegalovirus enhancer/promoter, a  
 CC bovine growth hormone polyA sequence, and a kanamycin resistance  
 CC selectable marker. Polynucleotides encoding polypeptides, such as  
 CC erythropoietin or leptin, and ribozymes and antisense  
 CC polynucleotides can be inserted into the vector. The vector is  
 CC preferred for use in novel compositions and methods for improved  
 CC polynucleotide delivery into cells. In these methods, polyclonetic  
 CC agents are used to increase the frequency of uptake of a  
 CC nucleic acid (see also V21684-86) into a cell. The polyclonetic  
 CC agent can condense with the nucleic acid and inhibit serum and/or  
 CC nuclease degradation of the nucleic acid. The nucleic acid can be  
 CC a vector, may express a therapeutic protein or a vaccinating viral  
 CC or cancer antigen, or is itself therapeutic (antisense or  
 CC ribozyme). The methods and compositions can be used in the gene  
 CC therapy of many diseases.  
 SQ Sequence 9600 BP: 2326 A: 2376 C: 2817 G: 2081 T:

Query Match 6.2%; Score 151.6; DB 1; Length 9600;  
 Best Local Similarity 53.2%; Pred. No. 6.8e-21;  
 Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

QY 1500 CTGTGCCCCCAGACGCTCTCCAGAGCGTGGCCCAACGGCTGCCCCAA 1559  
 DB 1333 CTCTGCCCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1274  
 QY 1560 CTGCGGACCTTCACCGCTGCTCCACGACGCGACCGCACTGACGCTCTCT 1619  
 DB 1273 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1214  
 QY 1620 CACCTACACGCAATTCCTACCTGCGGCGCCACCGCTGCGCCCACTCTCT 1679  
 DB 1213 CTGCGCTCTCTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1154  
 QY 1680 CTGTGCCCCCGAATCCCAACCACTGTCACAGTACGCTCTCTCTCTCTCT 1739  
 DB 1153 CTCTGCCCCCTGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1094  
 QY 1740 CCGCTACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1796  
 DB 1093 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1034  
 QY 1796 --GCCGCCACGCTGCGCCCAACCGCTGCGCCACGCGACGCGCTCTCTCT 1853  
 DB 1033 CCGCTCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 974  
 QY 1854 AATATCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1913  
 DB 973 CTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 914  
 QY 1914 CCACACACAGTCACTGCTACACCACTGCTGCGCCCTACTACGCGACAGCTGCGCA 1973  
 DB 913 CT 854  
 QY 1974 ACACACAGTCACTGCTACACCACTGCTGCGCCCACTGCACTGCTGCTGCTG 2033  
 DB 853 CCGCT 794  
 QY 2034 ATGACCCCAACACGAGCTGCGCGGTAACCTACACAGGCGACGAGCTACACACCTG 2093

DB 793 CTGCCCCCTGCTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 734  
 QY 2094 AAGATGATGACATGACACCGCCCTCTCTCCCAAGACACCCATCAACCTTGGCT 2145  
 DB 733 CTCTGCTCTCTGCGCCCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 682

RESULT 5  
 V55830/C  
 ID V55830 standard; DNA; 795 BP.  
 AC V55830;  
 DT 18-NOV-1998 (first entry)  
 DE FIGA insert stabilising polypeptide encoding DNA.  
 KW Fusion protein; stabilising polypeptide; proteolytic degradation;  
 KW resistance; half-life; autoimmune disease; inflammation; nitro drug;  
 KW IkappaB regulator protein; inflammatory bowel disease; in vivo imaging;  
 KW nitroreductase protein; enzyme therapy; produg therapy; protease;  
 KW cancer; pathological condition; ss.  
 OS Epstein-Barr virus.  
 FH Key  
 FT Key  
 FT CDS  
 Location/Qualifiers  
 1..788  
 /\*tag= a  
 /product= "stabilising polypeptide"  
 PN W09822577-A1.  
 PD 28-MAY-1998.  
 PF 17-NOV-1997; IB1508.  
 PR 25-JUN-1997; US-048945.  
 PR 15-NOV-1996; US-030986.  
 PA (MASU/) MASUCCI M G.  
 PI Masucci MG;  
 DR WPI: 98-312463/27.  
 DR P-PSDB: W79128.  
 PT New fusion proteins resistant to proteolytic degradation -  
 PT comprising a core protein with a stabilising polypeptide comprising  
 PT a peptide sequence containing glycine repeats  
 PS Disclosure: Fig 3; 120pp; English.  
 CC This DNA encodes a stabilising polypeptide and is the FIGA insert of the  
 CC invention. The invention provides a method for increasing the resistance  
 CC of a core protein to proteolytic degradation that comprises linking or  
 CC inserting onto or into the core protein a stabilising polypeptide of  
 CC formula (Gly)aX(Gly)bY(Gly)cZn where Glya, Glyb, Glyc are 1-6  
 CC sequential Gly residues and X, Y, Z are Ala, Ser, Val, Ile, Leu, Met,  
 CC Phe, Pro or Thr and n can be anything between 1-66. X, Y and Z need not  
 CC be identical from n repeat to n repeat. Alternatively a nucleic acid  
 CC encoding the stabilising polypeptide can be linked onto or inserted into  
 CC a nucleic acid encoding a core protein. The fusion proteins of the  
 CC invention are more resistant to degradation by proteases and, thus, have  
 CC a longer half-life than the unfused core protein. The products can be  
 CC used for treating autoimmune diseases, cancer and inflammation. In  
 CC particular, the core protein may be an IkappaB regulator protein for the  
 CC treatment of inflammatory bowel disease, or a nitroreductase protein  
 CC which can activate nitro drugs in enzyme/prodrug therapy to treat cancer  
 CC or other pathological conditions. The fusion proteins can also be used in  
 CC diagnostic methods such as in vivo imaging.  
 SQ Sequence 795 BP: 200 A: 104 C: 478 G: 13 T:

Query Match 5.8%; Score 142.4; DB 1; Length 795;  
 Best Local Similarity 53.2%; Pred. No. 2.5e-19;  
 Matches 351; Conservative 0; Mismatches 301; Indels 8; Gaps 2;

QY 1500 CTGTGCCCCGACAGCTCTCCAAAGGCTGCGCCCAACGGCTGCCCCAA 1559  
 DB 659 CTCTGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600  
 QY 1560 CTGCGGACCTTCACCGCTGCTCCACCTGCAAGCGCACCGCACTGACGCGCTCT 1619  
 DB 599 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 540  
 QY 1620 CACCTACACGCAATTCCTACCTGCGGCGCCACCGCTGCGCCCACTGCTCTCT 1679  
 DB 539 CTGCGCTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480





PI Gut J, Leech J, Nelson RC, Petersen C;  
DR WPI: 98-159290/14.  
DR P-PSDB: W48299.  
PT Anti-Cryptosporidium antibody - used to develop products for  
PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium  
PT infections  
PS Claim 32: Pages 62-63, 89pp: English.  
CC The sequence is that encoding the gp900 antigen which may be used  
CC in the production of anti-Cryptosporidium antibodies. These can be  
CC used for the prophylaxis, treatment, inhibition or retardation of  
CC a Cryptosporidium infection in humans or in animals such as calves.  
CC They can also be used for the detection and diagnosis of related  
CC infections.  
SQ Sequence 5318 BP; 1944 A; 1153 C; 888 G; 1333 T;

Query Match 5.1%; Score 124.2; DB 1; Length 5318;  
Best Local Similarity 51.4%; Pred. No. 1.1e-15;  
Matches 343; Conservative 0; Mismatches 313; Indels 11; Gaps 2;

OY 1493 ACATCGACTGCTGCCCCGAGCTGCTCCAAAGGCTGCCCAAGGCTGCCCAAGGCT 1552  
DB 584 ACAACAACAACACTACTACTACTACTACTAGACAACAACAACAACAACAACAACA 643  
OY 1553 GCCCAACTGCGCACCTCCACCGTGTCCCACTGCAAGCGCCGCACTGCAGCC 1612  
DB 644 ACTACTACAACTACTACTACTACTACTAGACAACAACAACAACAACAACAACA 703  
OY 1613 CAGTCCCACTCAACCGCAATTCTCAGTCCGCGCCCGCCGCTGCCCGCCGCAAGCT 1672  
DB 704 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 763  
OY 1673 CCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1732  
DB 764 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 823  
OY 1733 ACCGAGGCCCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1792  
DB 824 ACCACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 883  
OY 1793 GGTGCCCCCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1852  
DB 884 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 937  
OY 1853 GAATCCCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1912  
DB 937 --AACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 994  
OY 1913 CCCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1972  
DB 995 ACAACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1054  
OY 1973 AACACACAGTCACTGACACCAAGTGTGCGCCCGCCGCTGCTGCTGCTGCTGCTGCA 2032  
DB 1055 ACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACTACT 1114  
OY 2033 CATGACCCCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 2092  
DB 1115 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1174  
OY 2092 -TGAAGATGATGACATGACCCCTCTCTGCCCAACAACAACAACAACAACAACA 2150  
DB 1175 ACCAAGAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 1234  
OY 2151 AATGCA 2157  
DB 1235 AGGTAA 1241

RESULT 8  
098470/c  
ID 098470 standard; cDNA: 2744 BP.  
AC 098470;

DT 01-MAY-1996 (first entry)  
DE MispI-containing plasmid pMIS1.  
KW Spider silk; repeat unit; consensus; minor ampullate silk protein;  
KW spider; MispI; orb web spider; dragline; ds.  
OS Nephila clavipes.  
FH Key Location/Qualifiers  
FT cds 183..2678  
FT misc\_difference 996..1037  
FT /tag= a  
FT /note= "represented as indeterminate, a highly  
FT compressed GC rich region which could  
FT not be sequenced"

PN W09525165-A1.  
PD 21-SEP-1995.  
PE 14-MAR-1995: U03139.  
PR 14-MAR-1994: US-209747.  
PA (UWY-) UNITV WYOMING.  
PI Colgin M, Lewis RV;  
DR WPI: 95-336970/43.  
DR P-PSDB: R80168.  
PT Polypeptide(s) comprising repeated unit amino acid sequences, also  
PT derived from minor ampullate spider silk proteins and used  
PT to form spider silk fibres  
PS Claim 12: Fig 1: 86pp: English.  
CC cDNA clone, pMIS1, encoding the orb web spider minor ampullate silk  
CC protein MispI has been identified and sequenced. Repeat unit peptides  
CC of MispI may form part of a larger polypeptide with an amino terminus  
CC (see R80184-85) and a carboxy terminus (see R80186-89). Repeat unit  
CC peptides make up spider silk proteins (spidroins) which in turn  
CC aggregate to form the silk fibres. Spider silk fibres have high  
CC tensile strength and significant elasticity. An isolated cDNA clone of  
CC a silk protein encoding sequence is of use to produce the protein at  
CC high yields using recombinant DNA technology.  
SQ Sequence 2744 BP; 486 A; 549 C; 1059 G; 608 T;

Query Match 4.6%; Score 113; DB 1; Length 2744;  
Best Local Similarity 50.9%; Pred. No. 1.4e-13;  
Matches 269; Conservative 0; Mismatches 260; Indels 0; Gaps 0;

OY 1494 CATCGACTGCTGCCCCGAGCTGCTCCAAAGGCTGCCCAAGGCTGCCCAAGGCTG 1553  
DB 2121 CAGCGGCTCAGCTCTCTGAGGCGCTCCAGACACACACTCTACCTGTAACCCAGCC 2062  
OY 1554 CCCCACTGCGGACCTCCACCGTGTCCCACTGCAAGCGCACCGCACTGCAGCC 1613  
DB 2061 CTCGGATCTGACAGCAGCTGACAGCAGCAGCAGCTCTGCTGCGCAAGCTATCAC 2002  
OY 1614 CAGTCCCACTTACACCGAATTCTACTCTGCGGCCCGCCAGCGTGGCCCGCAGCTC 1673  
DB 2001 CTGACACACGTAAGCTCCAGGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1942  
OY 1674 CTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1733  
DB 1941 CAGCGGCTCTGACAGCAGCAGCTCTGTAACCCGAGCGCTCCAGCTCTGCTGCTG 1882  
OY 1734 CCGCAGCCCTTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1793  
DB 1881 CAGCTGCGGAGCAGCAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1822  
OY 1794 CTGCCCCCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1853  
DB 1821 CAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1762  
OY 1854 AAATCCCACTACTGTCATCATCACCACTACTGCTGCTGCTGCTGCTGCTGCTGCT 1913  
DB 1761 CAGCACCAGCAGCTCTACCTGTAACCGCAGCAGCTCTGCTGCTGCTGCTGCTGCT 1702  
OY 1914 CCAACACCAAGTCACTGTAACCAACCAAGTGTGCTGCTGCTGCTGCTGCTGCTGCT 1973  
DB 1701 CACCAAGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1642

```
OY 1974 ACACGACGACTGCTGACACCCGCTGCTCCCGCAGCTCCGCC 2022
DB 1641 CTGCACCTGCCCGAGCGGACGACGCTCCGACGACGACCTTACC 1593

RESULT 9
ID T80043/C
AC T80043 standard; DNA; 40875 BP.
DE 03-NOV-1997 (first entry)
DE Insert from cosmid 109.
KM Pectinase: yeast: polygalacturonase; phytopathogen: vegetable spoilage;
KM fruit spoilage; food processing; fruit juice; cosmid: ss.
OS Saccharomyces cerevisiae.
FH Location/Qualifiers
FT 312..1943
FT /*tag= a
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homology to
FT S. cerevisiae allantoate permease (DAL5)"
FT 3622..7107
FT /*tag= b
FT /note= "indicated in the specification as an open reading
FT frame"
FT 3965..5188
FT /*tag= c
FT /note= "indicated in the specification as an open reading
FT frame"
FT 3461..4546
FT /*tag= d
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homologies to
FT pectinases"
FT 5151..6194
FT /*tag= e
FT /note= "indicated in the specification as an open reading
FT frame, with homology to S. pombe thiamine
FT repressed gene"
FT 5486..5845
FT /*tag= f
FT /note= "indicated in the specification as an open reading
FT frame"
FT 6430..7470
FT /*tag= h
FT /note= "indicated in the specification as an open reading
FT frame"
FT 8051..8917
FT /*tag= i
FT /note= "indicated in the specification as an open reading
FT frame"
FT 9656..10552
FT /*tag= j
FT /note= "indicated in the specification as an open reading
FT frame"
FT 9986..10364
FT /*tag= k
FT /note= "indicated in the specification as an open reading
FT frame"
FT 10127..10936
FT /*tag= l
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homology to the
FT S. cerevisiae mitochondrial genome maintenance
FT protein (MGMI01)"
FT 11161..11523
FT /*tag= m
FT /note= "indicated in the specification as an open reading
FT frame"
FT 13441..14517
FT cds
```

```
FT /*tag= n
FT /note= "indicated in the specification as an open reading
FT frame"
FT 14058..14360
FT /*tag= o
FT /note= "indicated in the specification as an open reading
FT frame"
FT 14989..16119
FT /*tag= p
FT /note= "indicated in the specification as an open reading
FT frame"
FT 16405..17619
FT /*tag= q
FT /note= "indicated in the specification as an open reading
FT frame"
FT 16491..17336
FT /*tag= r
FT /note= "indicated in the specification as an open reading
FT frame, encoding a protein with homology to the
FT S. cerevisiae ribosomal protein S7A (SCRPS7A)"
FT 18429..18797
FT /*tag= s
FT /note= "indicated in the specification as an open reading
FT frame"
FT 18749..21037
FT /*tag= t
FT /note= "indicated in the specification as an open reading
FT frame"
FT 21763..22488
FT /*tag= u
FT /note= "indicated in the specification as an open reading
FT frame"
FT 22540..23262
FT /*tag= v
FT /note= "indicated in the specification as an open reading
FT frame"
FT 23667..28607
FT /*tag= w
FT /note= "indicated in the specification as an open reading
FT frame"
FT W09704106-A2.
FT PD 06-FEB-1997. E03091.
FT PF 15-JUL-1996; EP-870137.
FT PR 22-DEC-1995; EP-870089.
FT RA 14-JUL-1995; EP-870089.
FT PA (INNO-) INNOGENETICS NV.
FT PI Scarcez T, Van Broekhoven A;
FT DR WPI: 97-132645/12.
FT PT New nucleic acid encoding Saccharomyces pectinase - useful in food
FT processing, also its regulatory sequences are useful for expression
FT of heterologous protein
FT PS Example 1, Fig 2: 65pp; English.
FT CC This sequence represents the insert from cosmid 109, containing
FT CC Saccharomyces cerevisiae strain FY1679 DNA. This sequence contained the
FT CC pectinase gene of the invention (see T80042). Pectinases are synthesised
FT CC by plants and a variety of microorganisms, the best characterised
FT CC pectinase being polygalacturonase. Pectinases play roles in the invasion
FT CC of plant tissues by phytopathogens, the spoilage of fruits and
FT CC vegetables, and also in their food processing and plant biological
FT CC applications. Vectors containing the pectinase coding sequence are used
FT CC for production of homologous or heterologous proteins in transformed host
FT CC cells. The pectinase sequence may include regulatory and/or coding
FT CC sequences, including the sequence encoding the pectinase signal peptide
FT CC (which can be used in any recombinant system). The pectinase is useful in
FT CC food processing (e.g. preparation of fruit juices), typical heterologous
FT CC proteins that can be made are (poly)peptides of bacterial, animal or
FT CC viral origin. Also, probes derived from the pectinase coding sequence can
FT CC be used to detect the coding sequence and primers for amplification are
FT CC used to isolate related genes.
FT SQ Sequence 40875 BP; 12953 A; 7579 C; 8292 G; 12051 T;
```

Query Match

4.4%; Score 107.6; DB 1; Length 40875;











GenCore version 4.5  
Copyright (c) 1993 - 1998 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: July 18, 1999, 05:08:34 ; Search time 2471.47 Seconds

(without alignments)  
3553.707 Million cell updates/sec

Title: US-09-294-663-1

Perfect score: 2455

Sequence: 1 GTACGTTAAGTGAAGAAAAA.....AAAAAAAAAAAAAAAAAAAA 2455

Scoring table: IDENTITY\_NWC

Searched: 808301 seqs, 1788773984 residues

Database :

GenEmbl:\*  
1: gb\_ba1:\*  
2: gb\_ba2:\*  
3: gb\_in:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_ph:\*  
7: gb\_pl1:\*  
8: gb\_pl2:\*  
9: gb\_pl2:\*  
10: gb\_pl1:\*  
11: gb\_pl2:\*  
12: gb\_pl3:\*  
13: gb\_ro:\*  
14: gb\_st:\*  
15: gb\_sy:\*  
16: gb\_un:\*  
17: gb\_vl:\*  
18: gb\_hcg:\*  
19: em\_ba:\*  
20: em\_fun:\*  
21: em\_hum1:\*  
22: em\_hum2:\*  
23: em\_in:\*  
24: em\_om:\*  
25: em\_or:\*  
26: em\_ov:\*  
27: em\_pat:\*  
28: em\_ph:\*  
29: em\_pl:\*  
30: em\_ro:\*  
31: em\_sy:\*  
32: em\_un:\*  
33: em\_vl:\*  
34: em\_hcg:\*  
35: em\_sts:\*  
36: gb\_ba1:\*  
37: gb\_ba2:\*  
38: gb\_pl1:\*  
39: gb\_pl2:\*  
40: gb\_pl1:\*  
41: gb\_pl2:\*  
42: gb\_pl3:\*  
43: gb\_sts:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
-----				

1	2455	100.0	2455	3	TNAF000605	AF000605 Trichopl
2	2354.4	95.9	2821	3	TNAF000606	AF000606 Trichopl
3	193	7.9	15720	11	HUMMUC2X	L21998 Homo sapien
4	193	7.9	15720	41	HUMMUC2X	L21998 Homo sapien
5	151.8	6.2	185994	18	AC0002042	AC0002042 *** SEQUE
6	151.6	6.2	10596	6	I25041	I25041 Sequence 15
7	151.6	6.2	10596	6	I30503	I30503 Sequence 15
8	151.6	6.2	5452	15	U02454	U02454 Cloning vec
9	151.6	6.2	10850	15	U02455	U02455 Cloning vec
10	151.6	6.2	10737	15	XXM02428	U02428 Cloning vec
11	151.6	6.2	172281	17	EBV	V01555 Epstein-Bar
12	151.6	6.2	184113	17	HS4B958RAJ	M80517 Epstein-Bar
13	151.6	6.2	1150	17	HS4ULIR3	J02079 Epstein-Bar
14	145.6	5.9	3170	17	D88733	D88733 Equine herp
15	140	5.7	3173	17	D88734	D88734 Equine herp
16	132.8	5.4	2070	5	BSU43200	U43200 Boreogadus
17	130	5.3	2479	13	RNU87960	U87960 Rattus norv
18	127	5.2	3941	13	RNLAR2	X83546 R.norvegicu
19	125.8	5.1	33359	3	CEP55B11	Z83318 Caenorhabdi
20	124.4	5.1	28291	3	CEP56H9	Z74473 Caenorhabdi
21	121.6	5.0	5719	13	MMCATS1	AF051726 Mus muscu
22	120.8	4.9	1046	5	PSGAF402	M63477 P.americanu
23	118.8	4.8	4635	12	HS46805	AJ006805 Homo sapi
24	118.8	4.8	4635	42	HS46805	AJ006805 Homo sapi
25	118.4	4.8	1720	10	HUMMUC2A	M74027 Human mucin
26	118.4	4.8	1720	40	HUMMUC2A	M74027 Human mucin
27	117.6	4.8	174428	12	AC005821	AC005821 Homo sapi
28	117.6	4.8	174428	42	AC005821	AC005821 Homo sapi
29	116.4	4.7	176966	18	AC005139	AC005139 *** SEQUE
30	116	4.7	1988	11	AC002494	AC002494 Human PCR
31	116	4.7	1988	41	AC002494	AC002494 Human PCR
32	115.2	4.7	3811	10	HUMIMCA	M94131 Human mucin
33	115.2	4.7	3811	40	HUMIMCA	M94131 Human mucin
34	114.8	4.7	186324	18	AC005505	AC005505 *** SEQUE
35	114	4.6	123110	12	AC005383	AC005383 Homo sapi
36	114	4.6	123110	42	AC005383	AC005383 Homo sapi
37	113.8	4.6	11343	13	MMDSPPG	AJ002141 Mus muscu
38	113.8	4.6	4420	13	U67916	U67916 Mus muscu
39	113	4.6	2793	6	AR009990	AR009990 Sequence
40	113	4.6	2793	6	I95875	I95875 Sequence 1
41	111.4	4.5	3060	3	AF027735	AF027735 Nephila c
42	111	4.5	48000	18	AC005930	AC005930 *** SEQUE
43	108.6	4.4	3489	17	KSU52064	U52064 Kaposi's sa
44	108.4	4.4	1569	10	GIBINVOL	M35447 Hylobates l
45	108.4	4.4	1569	40	GIBINVOL	M35447 Hylobates l

## ALIGNMENTS

RESULT 1  
TNAF000605 2455 bp mRNA INV 28-JUN-1997  
LOCUS TNAF000605  
DEFINITION Trichoplusia ni insect intestinal mucin ITM14 mRNA, complete cds.  
ACCESSION AF000605  
NID 92224918  
KEYWORDS  
SOURCE cabbage looper.  
ORGANISM Trichoplusia ni  
Eukaryota; Eukaryotes; Metazoa; Arthropoda;  
Tracheata; Hexapoda; Insecta; Pterygota; Lepidoptera; Noctuoidea;  
Noctuidae; Trichoplusia.  
REFERENCE  
1 (bases 1 to 2455)  
AUTHORS Wang, P. and Granados, R.R.  
TITLE Molecular cloning and sequencing of a novel invertebrate intestinal  
mucin cDNA  
JOURNAL J. Biol. Chem. 272 (26), 16663-16669 (1997)  
MEDLINE 97341213  
REFERENCE 2 (bases 1 to 2455)  
AUTHORS Wang, P. and Granados, R.R.  
TITLE Direct Submission  
JOURNAL Submitted (21-APR-1997) Boyce Thompson Institute, Cornell  
University, Ithaca, NY 14853, USA



FEATURES  
source Location/Qualifiers  
1..2455  
/organism="Trichoplusia ni"  
/db\_xref="taxon:7111"  
38..2404  
CDS  
/codon\_start=1  
/product="Insect intestinal mucin IM14"  
/db\_xref="PID:g2224919"  
/translation="MIKTLFLTLALGLVAARPEVSDAEKRNPALEHPDPDPAEONHL  
LPHEDCTKFEYCEYGLKFIARDCAGTEFEKSEACVONAAALAGLILPGPAPITQA  
PATQAQATPTTQAQATPTTQAQATPTTQAQATPTTQAQATPTTQAQATPTTQA  
PTTQAQATPTTQAQATPTTQAQATPTTQAQATPTTQAQATPTTQAQATPTTQA  
AATTPAATTPGPAATPAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPVAPV  
TEFORCEGLYFNRYVQCDSPANVEDGEISAPVTEGNEDEDIDGLDNGCPA  
NEIDWLLPHGNRCDKYOCVHGNLVERGAGHSEFELQOCDHIELVCTLPGES  
EEVDVEDACTGWCTPEPIEMEPLEPNCGPADSFIDHLLPHESCGOYLQCVHQTIA  
RPGCNHFSAPQSCSPYTAGOVCECDSDNOCSTAAPTAPTAAPTAAPTAAPTA  
AABSTVVPAPTPAPTPAPTPPTTAIPTPAPTAAPTAPTPPAAPSPPTTVVPTAAPT  
AAPTVAPEIPIVTSAPTAAPTAPTAAPTAPTAAPTAPTAAPTAPTAAPTAPTAAPT  
APNTVTPPTAPTAAPTAPTAAPTAPTAAPTAPTAAPTAPTAAPTAPTAAPTAPTA  
TTPEDDIDPLPMDPIPCVIEECNVLPMAHADDKWVCDDGNNOVLYVCSGQFNP  
TTKACDPCACNVGCVRSNIOMSESEGYOVFLPMKLDLIDEDLRQALNFEU"

BASE COUNT 592 a 961 c 479 g 423 t  
ORIGIN

Query Match 100.0%; Score 2455; DB 3; Length 2455;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 2455; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GTAACTTAAGTGAAGAATTAACAGGCAAGTAATGATAAAGACCTCTATTCTT 60  
DB 1 GTAACTTAAGTGAAGAATTAACAGGCAAGTAATGATAAAGACCTCTATTCTT 60  
QY 61 GACGGCCCTCGGGCTCGCGCGCGCGCTCGAAGTCAAGCGAGCGGGAAGAACC 120  
DB 61 GACGGCCCTCGGGCTCGCGCGCGCGCTCGAAGTCAAGCGAGCGGGAAGAACC 120  
QY 121 TCTCCAGAGCGGCAACCCAGACTGCTCCGCTGAGCAGCAGTGGCTCTGCTCACA 180  
DB 121 TCTCCAGAGCGGCAACCCAGACTGCTCCGCTGAGCAGCAGTGGCTCTGCTCACA 180  
QY 181 ATAGAGCTGCAACCAATTCTACTACTGTGAATATGCTCAAGTTCATGCAACGAGA 240  
DB 181 ATAGAGCTGCAACCAATTCTACTACTGTGAATATGCTCAAGTTCATGCAACGAGA 240  
QY 241 CTGTGCTCTGTAACCAAGTTCCTCGGCTCAGACTGTGTCAAGCGCGCTTATG 300  
DB 241 CTGTGCTCTGTAACCAAGTTCCTCGGCTCAGACTGTGTCAAGCGCGCTTATG 300  
QY 301 CGGATGCACTGCGAGGACCTCCAGCTGAGCAACCCAGGCGCCAGCAACACTCAGGC 360  
DB 301 CGGATGCACTGCGAGGACCTCCAGCTGAGCAACCCAGGCGCCAGCAACACTCAGGC 360  
QY 361 CCCAACAACCAACCAAGGCGCCCAACCAACTACTAGGCGCCCTACTACACCAAGGC 420  
DB 361 CCCAACAACCAACCAAGGCGCCCAACCAACTACTAGGCGCCCTACTACACCAAGGC 420  
QY 421 CCCAACAACCAACCAAGGCGCCCAACCAAGGCGCCCAACCAAGGCGCCCAAGGC 480  
DB 421 CCCAACAACCAACCAAGGCGCCCAACCAAGGCGCCCAACCAAGGCGCCCAAGGC 480  
QY 481 AACTACACTAGGCGCCCTACTACTACTAGGCGCCCAACCAAGGCGCCCAAGGC 540  
DB 481 AACTACACTAGGCGCCCTACTACTACTAGGCGCCCAACCAAGGCGCCCAAGGC 540  
QY 541 TACACAACCAACCAAGGCGCCCAACCAAGGCGCCCAACCAAGGCGCCCAAGGC 600  
DB 541 TACACAACCAACCAAGGCGCCCAACCAAGGCGCCCAACCAAGGCGCCCAAGGC 600  
QY 601 TACACACTAGGCGCCCAACCAAGGCGCCCAACCAAGGCGCCCAAGGC 660  
DB 601 TACACACTAGGCGCCCAACCAAGGCGCCCAACCAAGGCGCCCAAGGC 660

QY 661 GGCCGCAACTACCCCGGCGCACTACCCCTGCGCGCAACCCCGCTGCAACTACCC 720  
DB 661 GGCCGCAACTACCCCGGCGCACTACCCCTGCGCGCAACCCCGCTGCAACTACCC 720  
QY 721 AGGTGTTCTGCAACCACTTACGCCCCAGTCTGCCCCGATCTGTACTGTTGCCAA 780  
DB 721 AGGTGTTCTGCAACCACTTACGCCCCAGTCTGCCCCGATCTGTACTGTTGCCAA 780  
QY 781 TGTGTGCGCACTGACCTTCGACATTCACCTGTTGATTCGCCAGACAACTGCACT 840  
DB 781 TGTGTGCGCACTGACCTTCGACATTCACCTGTTGATTCGCCAGACAACTGCACT 840  
QY 841 CTCTACAGAGTCTCCACAGGTTACACTTTCGAAACAGAGTGGCCCTAGGAGACTCTT 900  
DB 841 CTCTACAGAGTCTCCACAGGTTACACTTTCGAAACAGAGTGGCCCTAGGAGACTCTT 900  
QY 901 CAACCCCTACGTCACAGGCTGCGACTCTCTGCTAAGCTTGAATGCAAGCGCAATCAG 960  
DB 901 CAACCCCTACGTCACAGGCTGCGACTCTCTGCTAAGCTTGAATGCAAGCGCAATCAG 960  
QY 961 CCGCGCAACCCCGCACTGCAAGGCAAGCAAGAGCAATGCAATGCGAGACTCTCT 1020  
DB 961 CCGCGCAACCCCGCACTGCAAGGCAAGCAAGAGCAATGCAATGCGAGACTCTCT 1020  
QY 1021 CGACAAATGATGCGCCAGCTAATCTGAAATGCACTGCTTGGCCCAAGCAACCGTTG 1080  
DB 1021 CGACAAATGATGCGCCAGCTAATCTGAAATGCACTGCTTGGCCCAAGCAACCGTTG 1080  
QY 1081 CGACAAATGATGCGCCAGCTAATCTGAAATGCACTGCTTGGCCCAAGCAACCGTTG 1140  
DB 1081 CGACAAATGATGCGCCAGCTAATCTGAAATGCACTGCTTGGCCCAAGCAACCGTTG 1140  
QY 1141 CCACCTGATGTTGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1200  
DB 1141 CCACCTGATGTTGAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 1200  
QY 1201 CGGGGGGAGAGCGAAGAAAGTTGACGTGACGAGAGATGCTGACCGGCTGTACTGCC 1260  
DB 1201 CGGGGGGAGAGCGAAGAAAGTTGACGTGACGAGAGATGCTGACCGGCTGTACTGCC 1260  
QY 1261 CACGAAACCACTTAATGAGAGGCGCCCTCCCAAGGCTGCTCCAGTTCACATCGA 1320  
DB 1261 CACGAAACCACTTAATGAGAGGCGCCCTCCCAAGGCTGCTCCAGTTCACATCGA 1320  
QY 1321 CCACCTGCTCCCGCAG 1380  
DB 1321 CCACCTGCTCCCGCAG 1380  
QY 1381 TATGCAAGACCTTGCCTGGAACCTGCACTTCACTGCTGCAAGAGTCTGTGATC 1440  
DB 1381 TATGCAAGACCTTGCCTGGAACCTGCACTTCACTGCTGCAAGAGTCTGTGATC 1440  
QY 1441 TCTGTGACGCGCTGTGTTGCCAAGTTTTCGAGTGGGATTTCTGACACCAAGTGCAT 1500  
DB 1441 TCTGTGACGCGCTGTGTTGCCAAGTTTTCGAGTGGGATTTCTGACACCAAGTGCAT 1500  
QY 1501 TGTGTGCGCGAGAGCTGCTGCAAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
DB 1501 TGTGTGCGCGAGAGCTGCTGCAAGGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 1560  
QY 1561 TGGCGCAACCTTGCCTGAGTTCACCTGCAACGCGCAACCGCAACTGACAGCCCAAGTCC 1620  
DB 1561 TGGCGCAACCTTGCCTGAGTTCACCTGCAACGCGCAACCGCAACTGACAGCCCAAGTCC 1620  
QY 1621 ACTTACAAACGCAATTCCTACTCGGCGCCCAACCGCTGCGCCCAACCGCAAGTCTACTAC 1680  
DB 1621 ACTTACAAACGCAATTCCTACTCGGCGCCCAACCGCTGCGCCCAACCGCAAGTCTACTAC 1680  
QY 1681 TGTGTGCGCGAGAGTTCCTGCAAGGCTGCAAGGCTGCAAGGCTGCAAGGCTGCAAGG 1740  
DB 1681 TGTGTGCGCGAGAGTTCCTGCAAGGCTGCAAGGCTGCAAGGCTGCAAGGCTGCAAGG 1740



673 CCCGCGCCGAACCTACCCCTGCGCGACAAACCCCGCTGCACCTACCCAGGTGTTCTGC 732  
661 CCCGCGCCGAACCTACCCCTGCGCGACAAACCCCGCTGCACCTACCCAGGTGTTCTGC 720  
733 ACCCACTTCAGCCCGCAGTGTGCGCCCGCATCTGTGACCTGTTGCCAATGGTCCGACG 792  
721 ACCCACTTCAGCCCGCAGTGTGCGCCCGCATCTGTGACCTGTTGCCAATGGTCCGACG 780  
793 TGACTTCAGCATCAGCTTGTGTTCCCGACAGCAAGTACTCAACCTTTTACCAAGT 852  
781 TGACTTCAGCATCAGCTTGTGTTCCCGACAGCAAGTACTCAACCTTTTACCAAGT 840  
853 CTCCAAGGTTACACCTTGTGTTCCCGACAGCAAGTACTCAACCTTTTACCAAGT 912  
841 CTCGAAGGTTACACCTTGTGTTCCCGACAGCAAGTACTCAACCTTTTACCAAGT 900  
913 CCAGCGCTGCGACCTCTCTGCTTAACGTTGAATGCGACGCGCAATACAGCCCGACCCG 972  
901 CCAGCGCTGCGACCTCTCTGCTTAACGTTGAATGCGACGCGCAATACAGCCCGACCCG 960  
973 AGTCAGAGGCAAGCAAGAGCAAGACATGACATCGGAGACCTCTGCAATGAGATG 1032  
961 AGTCAGAGGCAAGCAAGAGCAAGACATGACATCGGAGACCTCTGCAATGAGATG 1020  
1033 CCCAGCTTAACCTTGAATGACCTGCTTTGCCCAAGGAAACCGTTGCGACATGATTA 1092  
1021 CCCAGCTTAACCTTGAATGACCTGCTTTGCCCAAGGAAACCGTTGCGACATGATTA 1080  
1093 CCAGTGTGCTCCAGGTAACCTTGTGAGAGGCGTTGTGAGAGCGCGACCCACTTCAATT 1152  
1081 CCAGTGTGCTCCAGGTAACCTTGTGAGAGGCGTTGTGAGAGCGCGACCCACTTCAATT 1140  
1153 TGAATTTAGCATGTGACACATGAGCTGTTGGCTGACCTCCCGGGGGGAGAG 1212  
1141 TGAATTTAGCATGTGACACATGAGCTGTTGGCTGACCTCCCGGGGGGAGAG 1200  
1213 CGAAGAAGTTGACGTGACGAGAGATGCTGACGCGCTGTGACTGCGCCACGAAACCAT 1272  
1201 CGAAGAAGTTGACGTGACGAGAGATGCTGACGCGCTGTGACTGCGCCACGAAACCAT 1260  
1273 TGAATGGAAGCCCTTCCCAACAGCGCTGCGCTGCGACTTACAGATGACCACTCTCTCC 1332  
1261 TGAATGGAAGCCCTTCCCAACAGCGCTGCGCTGCGACTTACAGATGACCACTCTCTCC 1320  
1333 CCACGAGAGCGACTGCGCGCATGATCTACAGTGTGCTCATGAGACAGATGAGCAAGAC 1392  
1321 CCACGAGAGCGACTGCGCGCATGATCTACAGTGTGCTCATGAGACAGATGAGCAAGAC 1380  
1393 TTGCCCCGAAAACTGACCTTCACTGCTGCGACACAGTCTGTGAGTCTCTGTGACCGC 1452  
1381 TTGCCCCGAAAACTGACCTTCACTGCTGCGACACAGTCTGTGAGTCTCTGTGACCGC 1440  
1453 TGGTTGCCAAGTTTTCAGTGTGATTTCTGACAACAAGTCAATGAGTCTGCTCCCGAC 1512  
1441 TGGTTGCCAAGTTTTCAGTGTGATTTCTGACAACAAGTCAATGAGTCTGCTCCCGAC 1500  
1513 AGCTGCTCAAGGGTGTGCGCCCAAGGCGTCCCAACGCGTGTGCGCCCAACGCGACCGC 1572  
1501 AGCTGCTCAAGGGTGTGCGCCCAAGGCGTCCCAACGCGTGTGCGCCCAACGCGACCGC 1560  
1573 CACCGTGTGCTGACCTGCAACGCGCAACCGCAACTGACGCGCCCAAGTCTCAACACCGC 1632  
1561 CACCGTGTGCTGACCTGCAACGCGCAACCGCAACTGACGCGCCCAAGTCTCAACACCGC 1620  
1633 AATTCCTACTCGCGCCCGCCAGCGCTGCGCCCAACGCGAGTCTCTACTACTGCTGCCCTGA 1692  
1621 AATTCCTACTCGCGCCCGCCAGCGCTGCGCCCAACGCGAGTCTCTACTACTGCTGCCCTGA 1680  
1693 ATTCGCAACCACTGTCAGTACAGTACAGTCTGCTGCTCCACGCGAGCCGCTACTACTGCT 1752  
1681 ATTCGCAACCACTGTCAGTACAGTACAGTCTGCTGCTCCACGCGAGCCGCTACTACTGCT 1740

1753 TGTCCCTGAATTCCTCACTGTCATGAGCGCTACCGCTGCTCCCGACCGCTGCC 1812  
1741 TGTCCCTGAATTCCTCACTGTCATGAGCGCTACCGCTGCTCCCGACCGCTGCC 1800  
1813 CACGCGTCCCGCCAGCGAGCGCTACTAGTGTGCTCCCAAAATCCCACTACTGTAC 1872  
1801 CACGCGTCCCGCCAGCGAGCGCTACTAGTGTGCTCCCAAAATCCCACTACTGTAC 1860  
1873 ATCAACCACTACTGCTGCGCCCACTAGCGAGCACTGCTCCCGCAACACAGTACTGT 1932  
1861 ATCAACCACTACTGCTGCGCCCACTAGCGAGCACTGCTCCCGCAACACAGTACTGT 1920  
1933 ACCACCACTGCT----- 1946  
1921 ACCACCACTGCTGCTCCCTACTACTAGCGAGCACTGCTCCCGCAACACAGTACTGAC 1980  
1946 -----GCCCTACTACTAGCGAGCACTGCTCCCGCAACACAGTACTGAC 1995  
1981 ACCACCGAGCGCTCTACTAGCGAGCACTGCTCCCGCAACACAGTACTGAC 2040  
1996 CACTGCTGCGCCCACTGAGCTCCCGCTACCGCTGCGACATGCAACACAGTACTG 2055  
2041 CACTGCTGCGCCCACTGAGCTCCCGCTACCGCTGCGACATGCAACACAGTACTG 2100  
2056 CCGGTTAATCAACACAGCGACAGTACTACACACCTGAGATGATGATGACATGAC 2115  
2101 CCGGTTAATCAACACAGCGACAGTACTACACACCTGAGATGATGATGACATGAC 2160  
2116 TCTCCCAAGCAACCCCACTGCTGCTGAGAAATGCAAGTGTGCTCCCGGCTCA 2175  
2161 TCTCCCAAGCAACCCCACTGCTGCTGAGAAATGCAAGTGTGCTCCCGGCTCA 2220  
2176 CGTACTGCGCAAAATCTGCTGCTGAGCGCAACCAAGTCTGCTGCTGCTGCTC 2235  
2221 CGTACTGCGCAAAATCTGCTGCTGAGCGCAACCAAGTCTGCTGCTGCTGCTC 2280  
2281 TGAGGCTTCCAGTTCACACCCCACTACTAAGACCTGTAATGCTGCTGCTGCTGCTG 2340  
2286 CGTGAAGCAACATTCAGTGTGAGAAAGTACGAAGATCCAGTCTTATCCCATG 2355  
2341 CGTGAAGCAACATTCAGTGTGAGAAAGTACGAAGATCCAGTCTTATCCCATG 2400  
2356 GAACAACATGATGAGATCAGACAGCGCTGAACTTGTGATGATGATGATGATGAT 2415  
2401 GAACAACATGATGAGATCAGACAGCGCTGAACTTGTGATGATGATGATGATGAT 2460  
2416 TTAATGAAGTTTGTGTTTA 2435  
2461 TTAATGAAGTTTGTGTTTA 2480

RESULT 3  
LOCUS HUMMUC2X 15720 bp mRNA PRI 09-JAN-1998  
DEFINITION Homo sapiens intestinal mucin (MUC2) mRNA, complete cds.  
ACCESSION L21998  
KEYWORDS g454153  
SOURCE human.  
ORGANISM Homo sapiens  
Eukaryotes; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;  
Primates; Catarrhini; Homnidae; Homo.  
REFERENCE 1 (bases 1 to 15720)  
AUTHORS Gun,J.R., Jr., Hicks,J.W., Toribara,N.W., Kim,Y.S. and Siddiki,B.  
TITLE Molecular cloning of human intestinal mucin (MUC2) cDNA:  
Identification of the amino terminal and overall sequence  
similarity to pre-pro-von Willebrand factor  
JOURNAL J Biol. Chem. 269, 2440-2446 (1994)  
MEDLINE 94132002  
REFERENCE 2 (bases 1 to 15720)  
AUTHORS Eckhardt,A.E., Timpler,C.S., Deluca,A.W. and Hill,R.L.



Db 5928 GACACCCATCACACCC---ACACATACGGTGTGACCCCAACCCCAACACCCACCCAGGACACACA 5984  
Oy 616 AACTACATCAACCCAGGCTGGAACACTACCCCGGCGCACTACCTCCCGGGCGCACTAACCC 675  
Db 5985 GACCCCAACGACGACACCCATACCAACCACTAGCGTGACCCCAACCCCAACCCAC 6044  
Oy 676 GGGCGCAATACCCCTGCGGACACACCCCGCTGCACTACCCAGGTGTCTCTGCACC 735  
Db 6045 GGGACACAGACCCCAACGACACCCATACCAACCACTAGCGTGACCCCAACCCCAACCC 6104  
Oy 736 CACTTACAGCCCACTGTCTGCGCCGATCTGTGACTGTGTCCCAATGGTGGCCAG---- 792  
Db 6105 AACACCCACCGGACACAGACCCCAACCAACCACTACACCCACCACTACCGTAC 6164  
Oy 792 -CTGACTTCGACATCACTGTGTGATTTGCCACGACAGTATGCACTCTCTTCACAG 850  
Db 6165 CCCAACCCCAACCCCAACCCGACACAGACCCCAACCAACCACTACCAACCAACCAAC 6224  
Oy 851 TGCTCCAAAGGTTACACCTTGGACAGAGTGCCCTGAGGAGCACTTCAACCCCTAC 910  
Db 6225 TACGGTACCCCAACCCCAACCCGACACAGAC---CCCAACGACGACACCAT 6281  
Oy 911 GTCCAGCGCTGGAGCTCTCTGCTAGCTTGAATGCAAGGCGAAATAGCCCGGACCC 970  
Db 6282 CACACCAACCACTAGCGTGACCCCAACCCCAACCCCAACCCGACACAGACCCCAAC 6341  
Oy 971 CCAGTCACAGAGGACGAGAGAGACATTTGACTGAGGAGGCTTGTGAAGCGGACCACTTCA 1030  
Db 6342 GACACCCATCAACCAACCACTAGCGTGACCCCAACCCCAACCCCAACCCGACACAG 6401  
Oy 1031 TGCCAGCTTAATCTGGAATGACTGCTTGTGCCCAAGAAACGTTGGACAAATAT 1090  
Db 6402 CCCAACCAAGACACCATACACACACCAACCAACCACTAGCGTGACCCCAACCCGAC 6461  
Oy 1091 TACCAGTCTGTCAGGTAATCTTGTAGAGAGGCTTGTGAAGCGGACCACTTCA 1150  
Db 6462 CACACAGACCCCAACCAACCAACCACTAGCGTGACCCCAACCCCAACCCGAC 6521  
Oy 1151 TTTGAACTTCACTGACATGACACATGACATGCTGTGGCTGACCCCTCGGGGCGGAG 1210  
Db 6522 ACCCAGCGGACACAGACCCCAACCAACCACTAGCGTGACCCCAACCCGAC 6581  
Oy 1211 AGCGAAGATTGACGTGACGAGAGATGCTGACCCGCTGTACTGCCCAACGAGACCC 1270  
Db 6582 AACCCCAACCAACCCGACACAGACCCCAACCAACCACTAGCGTGACCCCAACCCG 6641  
Oy 1271 ATTGAATGGAGCCCTCTCCCAACGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1330  
Db 6642 GGTGACCCCAACCCCAACCAACCCGACACAGACCCCAACCACTAGCGTGACCCCA 6701  
Oy 1331 CCCCGAGAGAGCTGCGGCAATGATATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1390  
Db 6702 CACCACTAGCGTGACCCCAACCCCAACCCGACACAGACCC---CCAACCAACGACA 6759  
Oy 1391 CTTGACCTGGAACCTGCACTTCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1450  
Db 6760 CCATACACACACACATAGCGTGACCCCAACCCCAACCCGACACAGACCCCA 6819  
Oy 1451 GCTGTTGCCAAGTTTGTGAGTGATTTGACAAACAGTGCATGCACTGCTGCTGCTGCT 1510  
Db 6820 ACCAGACACCCATACACACCACTAGCGTGACCCCAACCCCAACCCGACAC 6879  
Oy 1511 ACAGTGTCTCAAGGGTCCCAACGCGGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1570  
Db 6880 CAGACCCCAACCAACCACTACACCACTAGCGTGACCCCAACCCCAACCCGAC 6939  
Oy 1571 TCACAGCTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1624  
Db 6940 ACCGCAACAGACCCCAACCAACCACTAGCGTGACCCCAACCCCAACCCGAC 6999  
Oy 1625 ACAACCGCAATTTCTACTTCGCGGCGGACCGCTGCGGCGGACCGCTCTACTACTGCT 1684

Db 7000 CCAACACCCAGCGGACACAGACCCCAACCAACCACTAGCGTG 7059  
Oy 1685 GCGCTGAATCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCCCT 1744  
Db 7060 ACCCAACCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCCCT 7119  
Oy 1745 ACTACTGCTGTCTGGAATCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCAC 1804  
Db 7120 ACTAGGTCACCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCC 7179  
Oy 1805 GCTGCGGCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCGGACCG 1864  
Db 7180 ACCACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCGGACCGGAC 7239  
Oy 1865 ACTGTCACATCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCGGAC 1924  
Db 7240 ACACCCATACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCGGAC 7299  
Oy 1925 GTCACTGACACCCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1984  
Db 7300 CCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCGGACCGGAC 7359  
Oy 1985 ACTGTCACACCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2044  
Db 7360 ACACGACCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGACCG 7419  
Oy 2045 ACCAGAGTGGCGGTAATCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCAC 2102  
Db 7420 CCCACCGGACACAGACCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCAC 7479  
Oy 2103 ACATGACCCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2137  
Db 7480 ACCCAACACCCCAACCACTAGCGTGACAGTACCACTTGTGTGCTCCACCGGAC 7514

RESULT 4  
HUMMUC2X  
LOCUS HUMMUC2X 15720 bp mRNA PRI 09-JAN-1998  
DEFINITION Homo sapiens intestinal mucin (MUC2) mRNA, complete cds.  
ACCESSION L21998  
NID g454153  
KEYWORDS  
SOURCE human.  
ORGANISM human.  
REFERENCE  
AUTHORS Gum, J.R., Jr., Hicks, J.W., Toribara, N.W., Kim, Y.S. and Siddiki, B.  
TITLE Molecular cloning of human intestinal mucin (MUC2) cDNA: identification of the amino terminal and overall sequence similarity to pre-pro-von Willebrand factor  
JOURNAL J. Biol. Chem. 269, 2440-2446 (1994)  
MEDLINE 94132002  
REFERENCE  
AUTHORS Eckhardt, A.E., Timpte, C.S., Deluca, A.W. and Hill, R.L.  
TITLE The complete cDNA sequence and structural polymorphism of the polypeptide chain of porcine submaxillary mucin  
JOURNAL J. Biol. Chem. 272 (52), 33204-33210 (1997)  
MEDLINE 98070526  
FEATURES  
source  
1..15720  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/tissue\_type="intestine"  
28..15567  
/codon\_start=1  
/product="mucin"  
/translation="MGIPLARLAVALCAISLAGSEIOTEGRTRYHGRNYCSWGNPH  
YKFDGIVRFEPGLCDYNFASDGRSITKRSYAVLKRPGGAEPAGVESILITIKDDT  
IYLRHLAVNGAVVSTPHYSPLILKESDAVTKVSRALTLMMNEDLMLKIDTK  
FRNHTGICDYNGLQSYSEFLDVFLEFSLKNGKINQPVVEDPEEVAAPASC  
SEHRAECERILLTAFAFADCDLVPLEPYLRACQODRCRPGGDTVCSTVAEFSRCS

[illegible][illegible]



```

QY 851 TGCTCAAGGTTACACCTTGGACAGAGGTGCTTGGAGGACTCTACTTCAACCCCTAC 910
    || || || || || || || || || || || || || || || || || || || || ||
Db 6225 TAGGGTAGCCCCAACCCCAACACCCACCGGACACAGAC---CCACACGACAGCCCAT 6281
QY 911 GTCCAGCGCTCGAGCTCTCTGCTAAGCTTGAAAGCGAGCGGAATGACCCCGCACCC 970
    || || || || || || || || || || || || || || || || || || || || ||
Db 6282 CACCAACCCACCTAGAGGTAGCCCCCAACCCCAACCCCGGACACAGACCCCAACCC 6341
QY 971 CCAGTCACAGAGGACAGAGAGACATGACATCGGAGACTCTCTGACCAATGGA 1030
    || || || || || || || || || || || || || || || || || || || || ||
Db 6342 GACACCCATCACACACACCTAGCGTGACCCCAACCCCAACCCCGGACACAGAC 6401
QY 1031 TGCCAGCTTAATCTGGAATGAGCTGTGCCCCAGGAACCGTTGGACAAATAT 1090
    || || || || || || || || || || || || || || || || || || || || ||
Db 6402 CCCAACACGACACCCATACACACACGACACCTAGGTAACCCCAACCCCAACCCG 6461
QY 1091 TACAGCTCGTCACAGGTAATCTGTAGAGAGCGTGTGAGCCGCACTTCAGT 1150
    || || || || || || || || || || || || || || || || || || || || ||
Db 6462 CACACAGACCCCAACCCACACCCATCACACACCTAGGTAACCCCAACCCCAAC 6521
QY 1151 TTGAACCTTACAGATGACATGACATGAGCTGTGCTGACCTCCCGGGGCGAG 1210
    || || || || || || || || || || || || || || || || || || || || ||
Db 6522 ACCCAGCGGACACGACCCCAACCCACACACCTACACCTAGGTAACCC 6581
QY 1211 AGCGAAGAGTTGAGCTGACGAGAGATGCTGACCCGCTGTACTGCCACAGAACCC 1270
    || || || || || || || || || || || || || || || || || || || || ||
Db 6582 AACCCCAACACCCACCGGACACACACCCCAACCCCAACCCCAACCCCAACCCAC 6641
QY 1271 ATTGAAATGGAGCCCTCCCAACCGGCTGCTGCGGACTTCAGATGACACCTCTC 1330
    || || || || || || || || || || || || || || || || || || || || ||
Db 6642 GGTGACCCCAACCCCAACCCGACACAGAGACCCCAACCCCAACCCCAACCCATCAC 6701
QY 1331 CCCACGAGAGAGCTGGGGCAGTATACAGTGTGACATGAGACAGATATGCAAGA 1390
    || || || || || || || || || || || || || || || || || || || || ||
Db 6702 CACCACTAGGTAACCCCAACCCCAACCCGACACAGACC--CCAACCAAGACA 6759
QY 1391 CTTGCCCCGGAACCTGCACTTCACTGCTGACACAGTCTGTGAGTCTCTTGACC 1450
    || || || || || || || || || || || || || || || || || || || || ||
Db 6760 CCATACACACGACACAGTACGTTGACCCCAACCCCAACCCGACACAGACCCCA 6819
QY 1451 GCTGTTGCCAAGTTTGTGAGTGTGACATGACACAGTGCATGCACTGCTGCCCG 1510
    || || || || || || || || || || || || || || || || || || || || ||
Db 6820 ACCACGACACCCATCACACACCACTAGGTAACCCCAACCCCAACCCGACACA 6879
QY 1511 ACAGTGTCTCAAGGGTGGCCCAAGGGTGGCCCAAGGGTGGCCCAAGGGTGGCC 1570
    || || || || || || || || || || || || || || || || || || || || ||
Db 6880 CAGACCCCAACGACACACCACTACACACACCACTAGGTAACCCCAACCCCAACCC 6939
QY 1571 TCACACCTGTCACACCTGACACGACCCGCACTGAGCCCT---AGTCCACCT 1624
    || || || || || || || || || || || || || || || || || || || || ||
Db 6940 ACCGACACACAGACCCCAACCCGACACCACTACACCACTAGGTAACCCCAACCC 6999
QY 1625 ACAACCGCAATCTCTACTCGGGGCCCCACCGTGGCCCCACCGCTCTACTACTGCT 1684
    || || || || || || || || || || || || || || || || || || || || ||
Db 7000 CCACACCCACCGGACACAGACCAACCAACCACTACCACTACCACTACGTTG 7059
QY 1685 GCGCCTGATCCCAACGACGTGACAGTACACCTACTGCTGCTCCACCGCACCCCT 1744
    || || || || || || || || || || || || || || || || || || || || ||
Db 7060 ACCCAACCCCAACGACCCGACACAGACCCCAACCCCAACCCCAACCCCAACCC 7119
QY 1745 ACTACTGCTGCTGGAATCCCAATCTGACATGACAGCTGACCTGCTGCTGCTGCT 1804
    || || || || || || || || || || || || || || || || || || || || ||
Db 7120 ACTAGGAGACCCCAACCCCAACCCGACACAGACCAACCCCAACCCCAACCCCAAC 7179
QY 1805 GCTGCCCCACCGCTGCCCCACCGACCGCTACTACTGCTGCTGCTGCTGCTGCT 1864
    || || || || || || || || || || || || || || || || || || || || ||
Db 7180 ACCACCACTACTAGGTAACCCCAACCCCAACCCGACACAGACCCCAACCCCAAC 7239
QY 1865 ACTGTCATGACACGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1924
    || || || || || || || || || || || || || || || || || || || || ||
Db 7240 ACACCACTACTAGGTAACCCCAACCCCAACCCGACACAGACCCCAACCCCAACCC 7299

```

```

QY 1925 GTCAGTATCCACCCACTGCTGCTGCTACTACCCGACGACTGCCCCCAACCCAGATC 1984
    || || || || || || || || || || || || || || || || || || || || ||
Db 7300 CCAACACGACACCCATACACACCACTAGGTAACCCCAACCCCAACCCCAACCCGAC 7359
QY 1985 ACTGATCCACCCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2044
    || || || || || || || || || || || || || || || || || || || || ||
Db 7360 ACAGAGACCCCAACGACACCCATACACACCACTAGGTAACCCCAACCCCAACCA 7419
QY 2045 ACCAGAGTGGCCCGTACTAGTACCAACGAGGACGACGACGACGACGACGACGAC 2102
    || || || || || || || || || || || || || || || || || || || || ||
Db 7420 CCCACCGGACACAGACCCCAACCCCAACCCCAACCCCAACCCCAACCCCAACCC 7479
QY 2103 ACATGACCCCTCTCTCCCAACGACCCCAAC 2137
    || || || || || || || || || || || || || || || || || || || || ||
Db 7480 ACCCAACACCCGACGACAGACCCCAACCCCAAC 7514

```

```

RESULT 5
AC002042/c DNA HTG 22-APR-1998
LOCUS *** SEQUENCING IN PROGRESS *** Homo sapiens chromosome 16p11.2 BAC
DEFINITION clone C1987SK-A-18062; HTGS phase 1, 5 unordered pieces.
ACCESSION AC002042
NID g3075381
KEYWORDS HTG; HTGS_PHASE1.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Mammalia; Eutheria;
Primates; Catarrhini; Hominoidea; Homo.
REFERENCE 1 (bases 1 to 185994)
AUTHORS Adams,M.D., Loftus,B.J., Zhou,L., Phillips,C., Brandon,R.C.,
Fuhrmann,J., Kim,U.J., Kerlavage,A.R. and Venter,J.C.
Human chromosome 16p13 BAC clone C1987SK-A-18062
Unpublished
2 (bases 1 to 185994)
Adams,M.D. and Loftus,B.J.
Direct Submission
Submitted (29-APR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Apr 22, 1998 this sequence version replaced gi:3068565.
***
*** WARNING: Phase 1 High Throughput Genome Sequence ***
***
* This sequence is unfinished. It consists of 5 contigs for
* which the order is not known; their order in this record is
* arbitrary. In some cases, the exact lengths of the gaps
* between the contigs are also unknown; these gaps are presented
* as runs of N as a convenience only. When sequencing is complete,
* the sequence data presented in this record will be replaced
* by a single finished sequence with the same accession number.
* 1 2198: contig of 2198 bp in length.
* 2 2199: contig of unknown length.
* 3 2249: gap of unknown length.
* 4 2248: gap of unknown length.
* 5 2249: gap of unknown length.
* 6 46288: contig of 44040 bp in length.
* 7 46289: gap of unknown length.
* 8 46338: gap of unknown length.
* 9 46339: contig of 64581 bp in length.
* 10 110920: gap of unknown length.
* 11 110920: contig of 44256 bp in length.
* 12 145225: gap of unknown length.
* 13 145276: contig of 40719 bp in length.
* 14 145276: contig of 40719 bp in length.

```

```

FEATURES
source 1..185994
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="A-18062"
BASE COUNT 51275 a 43334 c 43047 g 48137 t 201 others
ORIGIN

```

```

Query Match 6.2%: Score 151.8; DB 18; Length 185994;
Best Local Similarity 53.7%: Pred. No. 1.6e-19;
Matches 337; Conservative 0; Mismatches 287; Indels 3; Gaps 1;
QY 1497 CGACTGTGCTCCGACAGCTCTCAACGGCTGCCCAACGGCTGCCCAACGGCTGCC 1556

```

Db 111975 CTACTCTCTGCGACAGCTACTGCACTGCTGCTACTCTCCCTCCGACAGCTACTGCA 111916  
1557 CAAGTGGCGACCCCTCCAGCGTGGTCCACCTGCAAGCCGACAGCTGACAGCCGAG 1616  
Db 111915 CTGCTGCTGCTGCTACTCTCTCTCTGCGACAGCTACTGCACTACTGATGCTCTCTCTG 111856  
1617 TCCACCTACCAACCGCAATTCTACTCCGCGCCCGACCGCTGCGCCGACAGCTGCTA 1676  
Db 111855 CCACAGCTACTGCAACAGCTGATGCTACTCTCTCCGCGACAGCTACTGCACTGCTG 111796  
1677 CTACTGCTGCGCCCTGAAATCCCAACAGCTGCAAGTACCACTACTGCTGCTCCACCG 1736  
Db 111795 ATGCTGCTACTCTCTCTGCGACAGCTACTGCACTGCTGATGCTGCTACTCTCTGCA 111736  
1737 CAGCCCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796  
Db 111735 CAGCTACTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111676  
1797 CCGCCACCGCTGCGCCCGACCGCTGCGCCCGACCGCCCTACTACTGCTGCTCCAGAA 1856  
Db 111675 CTGCTACTCTGCGACAGCTACTGCACTGCTGATGCTGCTGCTGCTGCTGCTGCTG 111616  
1857 TCCACCTACTGCTGCAATCAACAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1916  
Db 111615 CTGCACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111556  
1917 ACACAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1976  
Db 111555 CTCTCTCTGCGACAGCTACTGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111499  
1977 CCAAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2036  
Db 111498 CTGCACTGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111439  
2037 CACCCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2096  
Db 111438 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111379  
2097 ATGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2123  
Db 111378 CTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 111352

RESULT 6  
LOCUS 125041 10596 bp DNA PAT 21-AUG-1996  
DEFINITION Sequence 15 from patent US 5547856.  
ACCESSION 125041  
NID 91604911  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 10596)  
AUTHORS Godowski, P.J., Lokker, N.A. and Mark, M.R.  
TITLE Hepatocyte growth factor variants  
JOURNAL Patent: US 5547856-A 15 20-AUG-1996;  
FEATURES  
source 1.10596  
/organism="unknown"  
BASE COUNT 2627 a 2571 c 3023 g 2375 t  
ORIGIN

Query Match 6.2%; Score 151.6; DB 6; Length 10596;  
Best Local Similarity 53.2%; Pred. No. 2.2e-19;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

QY 1500 CTGCTGCGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559  
Db 2867 CTCTGCTGCGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808

QY 1560 CTGCGCGACCTCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619  
Db 2807 CTCTGCTGCGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2748  
1620 CACCTACACCGCAATTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1679  
Db 2747 CTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2688  
1680 CTGCTGCGCGCGAATCCCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1739  
Db 2687 CTCTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2628  
1740 CCGCTACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796  
Db 2627 CTCTGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2568  
1796 --GCCCGACCGCTGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1853  
Db 2567 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2508  
1854 AATCCCACTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913  
Db 2507 CTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2448  
1914 CCAACACAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1973  
Db 2447 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2388  
1974 ACACAGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2033  
Db 2387 CCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2328  
2034 ATGACACCAACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2093  
Db 2327 CTGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2268  
2094 AAGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2145  
Db 2267 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2216

RESULT 7  
LOCUS 130503 10596 bp DNA PAT 13-DEC-1996  
DEFINITION Sequence 15 from patent US 5580963.  
ACCESSION 130503  
NID 91821294  
KEYWORDS  
SOURCE Unknown.  
ORGANISM Unclassified.  
REFERENCE 1 (bases 1 to 10596)  
AUTHORS Godowski, P.J., Lokker, N.A. and Mark, M.R.  
TITLE Single-chain hepatocyte growth factor variants  
JOURNAL Patent: US 5580963-A 15 03-DEC-1996;  
FEATURES  
source 1.10596  
/organism="unknown"  
BASE COUNT 2627 a 2571 c 3023 g 2375 t  
ORIGIN

Query Match 6.2%; Score 151.6; DB 6; Length 10596;  
Best Local Similarity 53.2%; Pred. No. 2.2e-19;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

QY 1500 CTGCTGCGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1559  
Db 2867 CTCTGCTGCGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2808  
1560 CTGCGCGACCTCCACCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1619  
Db 2807 CTCTGCTGCGCGACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2748





```

REFERENCE          2 (bases 1 to 10850)
AUTHORS            Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
TITLE              CDNA expression cloning in human cells using the plambadr2
JOURNAL            Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
REFERENCE          3 (bases 1 to 10850)
AUTHORS            Kitz,P.A.
TITLE              Direct Submission
JOURNAL            Submitted (07-OCT-1993) Paul A. Kitz, CLONTECH Laboratories, Inc.,
                  1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT            Lambda DR2 can be obtained from CLONTECH Laboratories, Inc., 1020
                  East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
                  call (415) 424-8222 or (800) 662-2566, extension 1. International
                  customers, please contact your local distributor. For technical
                  information, call (415) 424-8222 or (800) 662-2566, extension 3.
                  This sequence was compiled by Andrew Murphy and revised at
                  CLONTECH. If you suspect there is an error in this sequence, please
                  contact CLONTECH's Technical Service Department at (415) 424-8222
                  or (800) 662-2566, extension 3 or E-mail TECH@CLONTECH.COM.
FEATURES
  source            1.10850
                   /organism="Cloning vector IPDR2"
                   /note="plasmid released from lambda DR2"
                   /db_xref="taxon:31858"
BASE COUNT        2723 a 2615 c 3053 g 2459 t
ORIGIN
Query Match       6.2%; Score 151.6; DB 15; Length 10850;
Best Local Similarity 53.2%; Pred. No. 2.2e-19;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

QY 1500 CTGCTGCCCCGAGCTGCTGCTCAACGCGTGCCTGCCCCCAACGCGTGCCTGCCCCCA 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7347 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7288

QY 1560 CTGCGCGACCTCTCCACCGTGTGCTCCACCTGCAACGCGACCGCGACCTGCAAGCTCC 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7287 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7228

QY 1620 CACTTCAACCGCAATTCCTACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACTA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7227 CTGCCCCCTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7168

QY 1680 CTGCTGCGCTGATGATCCCAACGCTGCTACATACACTACTGCTGCTGCTGCTGCTGCTG 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7167 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7108

QY 1740 CCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7107 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7048

QY 1796 --GCCCCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACTGCTGCTG 1853
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7047 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6988

QY 1854 AATGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1913
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6987 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6928

QY 1914 CCAACACAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1973
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6927 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6868

QY 1974 ACACACAGTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2033
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6867 CCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6808

QY 2034 ATGACACCAACAGCAGCGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2093
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6807 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6748

QY 2094 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2145

```

```

Db 6747 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6696
          ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
RESULT 10
LOCUS            XX002428/c
DEFINITION       Cloning vector pDR2, complete sequence.
ACCESSION        002428
KEYWORDS         9413794
SOURCE            Cloning vector pDR2.
ORGANISM          Cloning vector pDR2
REFERENCE         1 (bases 1 to 10737)
AUTHORS           Murphy,A.J., Kung,A.L., Swirski,R.A. and Schimke,R.T.
TITLE             CDNA expression cloning in human cells using the plambadr2
JOURNAL           Methods: A Companion to Methods in Enzymology 4, 111-131 (1992)
REFERENCE         2 (bases 1 to 10737)
AUTHORS           Kitz,P.A.
TITLE             Direct Submission
JOURNAL           Submitted (07-OCT-1993) Paul A. Kitz, CLONTECH Laboratories, Inc.,
                  1020 East Meadow Circle, Palo Alto, CA 94303, USA
COMMENT            This vector can be obtained from CLONTECH Laboratories, Inc., 1020
                  East Meadow Circle, Palo Alto, CA 94303, USA. To place an order
                  call (415) 424-8222 or (800) 662-2566, extension 1. International
                  customers, please contact your local distributor. For technical
                  information, call (415) 424-8222 or (800) 662-2566, extension 3.
                  This sequence has been compiled from information in the sequence
                  databases, published literature and other sources, together with
                  partial sequences obtained by CLONTECH: this vector has not been
                  completely sequenced. If you suspect there is an error in this
                  sequence, please contact CLONTECH's Technical Service Department at
                  (415) 424-8222 or (800) 662-2566, extension 3 or E-mail
                  TECH@CLONTECH.COM.
FEATURES
  source            1.10737
                   /organism="Cloning vector pDR2"
                   /db_xref="taxon:31805"
BASE COUNT        2686 a 2596 c 3033 g 2422 t
ORIGIN
Query Match       6.2%; Score 151.6; DB 15; Length 10737;
Best Local Similarity 53.2%; Pred. No. 2.2e-19;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

QY 1500 CTGCTGCCCCGAGCTGCTGCTCAACGCGTGCCTGCCCCCAACGCGTGCCTGCCCCCA 1559
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7234 CTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7175

QY 1560 CTGCGCGACCTCTCCACCGTGTGCTCCACCTGCAACGCGACCGCGACCTGCAAGCTCC 1619
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7174 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7115

QY 1620 CACTTCAACCGCAATTCCTACTGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACTA 1679
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7114 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 7055

QY 1680 CTGCTGCGCTGATGATCCCAACGCTGCTACATACACTACTGCTGCTGCTGCTGCTGCTG 1739
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 7054 CTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6995

QY 1740 CCTCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 6994 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 6935

QY 1796 --GCCCCACCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTACTGCTGCTG 1853

```

Db 6934 CCCCCTCCCTGCTCCGCCCCCTCTGCCCCCTCCCTGCTCTGCCCCCTC 6875  
1854 AATCCCAACTGTCACATACACACTACTGTCGCCCTACACGACGACCTGCCC 1913  
Db 6874 CTGCCCCCTCTGCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTCTGCTC 6815  
1914 CCAACACACAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 1973  
Db 6814 CTCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6755  
1974 ACACACAGTACTGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 2033  
Db 6754 CCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6695  
2034 ATGCACCAACACACAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAGTCAG 2093  
Db 6694 CTGCCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6635  
2094 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2145  
Db 6634 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 6583

## RESULT 11

EBV 172281 bp DNA circular VRL 17-AUG-1998  
LOCUS Epstein-Barr virus (EBV) genome. The complete sequence [1-10] was  
DEFINITION determined from DNA from B95-8 cells cloned by Arrand et al [11].  
B95-8 is a productive marmoset lymphoblastoid cell line  
immortalized with human EBV from a mononucleosis patient.  
V0155 J02070 K01729 K01730 V01554 X00498 X00499 X00784  
959074

KEYWORDS DNA polymerase; EBNA: genome; ribonucleotide reductase; tandem  
repeat; terminal repeat.  
SOURCE Epstein-Barr virus.  
ORGANISM Human herpesvirus 4  
Virus; dsDNA viruses, no RNA stage: Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE 1 (bases 1 to 172281)  
Arrand, J.R., Rymo, L., Walsh, J.E., Bjorck, E., Lindahl, T. and  
Griffin, B.E.

JOURNAL Molecular cloning of the complete Epstein-Barr virus genome as a  
MEDLINE set of overlapping restriction endonuclease fragments  
AUTHORS Nucleic Acids Res. 9 (13), 2999-3014 (1981)  
82014887

REFERENCE 2 (bases 1 to 172281)  
Kozak, M.

JOURNAL Possible role of flanking nucleotides in recognition of the AUG  
MEDLINE Initiator codon by eukaryotic ribosomes  
AUTHORS Nucleic Acids Res. 9 (20), 5233-5262 (1981)  
82059504

REFERENCE 3 (bases 1 to 172281)  
Deininger, P.L., Bankier, A., Farrell, P., Baer, R. and Barrell, B.  
AUTHORS Sequence analysis and in vitro transcription of portions of the  
TITLE Epstein-Barr virus genome  
J. Cell. Biochem. 19 (3), 267-274 (1982)  
83109311

JOURNAL 4 (bases 1 to 172281)  
MEDLINE Farrell, P.J., Deininger, P.L., Bankier, A. and Barrell, B.  
AUTHORS Homologous upstream sequences near Epstein-Barr virus promoters  
TITLE Proc. Natl. Acad. Sci. U.S.A. 80 (6), 1565-1569 (1983)  
83169725

REFERENCE 5 (bases 45644 to 52450)  
Jeang, K.T. and Hayward, S.D.  
AUTHORS Organization of the Epstein-Barr virus DNA molecule. III. Location  
TITLE of the P3HR-1 deletion junction and characterization of the NotI  
repeat units that form part of the template for an abundant  
13-O-tetradecanoylphorbol-13-acetate-induced mRNA transcript  
J. Virol. 48 (1), 135-148 (1983)  
83294686

JOURNAL 6 (bases 159853 to 172281)  
MEDLINE Bankier, A.T., Deininger, P.L., Satchwell, S.C., Baer, R., Farrell, P.J.  
REFERENCE AUTHORS

TITLE and Barrell, B.G.  
JOURNAL DNA sequence analysis of the EcoRI Dhet fragment of B95-8  
MEDLINE Epstein-Barr virus containing the terminal repeat sequences  
AUTHORS Mol. Biol. Med. 1 (4), 425-445 (1983)  
85060428

REFERENCE 7 (bases 1 to 172281)  
Farrell, P.J., Bankier, A.T., Seguin, C., Deininger, P.L. and  
Barrell, B.G.  
TITLE Latent and lytic cycle promoters of Epstein-Barr virus  
JOURNAL EMBO J. 2, 1331-1338 (1983)  
REFERENCE 8 (bases 142687 to 159853)  
Bankier, A.T., Deininger, P.L., Farrell, P.J. and Barrell, B.G.  
AUTHORS Sequence analysis of the 17,166 base-pair EcoRI fragment C of B95-8  
TITLE Epstein-Barr virus  
Mol. Biol. Med. 1 (1), 21-45 (1983)  
85035713

JOURNAL 9 (bases 112620 to 125316)  
MEDLINE Seguin, C., Farrell, P.J. and Barrell, B.G.  
REFERENCE DNA sequence and transcription of the BamHI fragment B region of  
AUTHORS B95-8 Epstein-Barr virus  
Mol. Biol. Med. 1 (3), 369-392 (1983)  
85060424

JOURNAL 10 (bases 45415 to 52824)  
MEDLINE Jones, M.D., Foster, L., Sneyd, T. and Griffin, B.E.  
REFERENCE The EB virus genome in Daudi Burkitt's lymphoma cells has a  
AUTHORS deletion similar to that observed in a non-transforming strain  
TITLE (P3HR-1) of the virus  
EMBO J. 3 (4), 813-821 (1984)  
84207939

JOURNAL 11 (bases 87650 to 92703)  
MEDLINE Biggin, M., Farrell, P.J. and Barrell, B.G.  
REFERENCE Transcription and DNA sequence of the BamHI L fragment of B95-8  
AUTHORS Epstein-Barr virus  
EMBO J. 3 (5), 1083-1090 (1984)  
84236104

JOURNAL 12 (bases 7315 to 9312)  
MEDLINE Yates, J., Warren, N., Reisman, D. and Sugden, B.  
REFERENCE A cis-acting element from the Epstein-Barr viral genome that  
AUTHORS permits stable replication of recombinant plasmids in latently  
TITLE infected cells  
Proc. Natl. Acad. Sci. U.S.A. 81 (12), 3806-3810 (1984)  
84222045

JOURNAL 13 (bases 76089 to 79808)  
MEDLINE Gibson, T., Stockwell, P., Ginsburg, M. and Barrell, B.  
REFERENCE Homology between two EBV early genes and HSV ribonucleotide  
AUTHORS reductase and 38k genes  
TITLE Nucleic Acids Res. 12 (12), 5087-5099 (1984)  
84247360

JOURNAL 14 (bases 1 to 172281)  
MEDLINE Baer, R., Bankier, A.T., Biggin, M.D., Deininger, P.L., Fallell, P.J.,  
AUTHORS Gibson, T.J., Hatfull, G., Hudson, G.S., Satchwell, S.C., Seguin, C.,  
TITLE Tuffnell, P.S. and Barrell, B.G.  
DNA sequence and expression of the B95-8 Epstein-Barr virus genome  
Nature 310 (5974), 207-211 (1984)  
84270667

JOURNAL 15 (bases 1 to 172281)  
MEDLINE Bodescot, M. and Perricaudet, M.  
REFERENCE Clustered alternative splice sites in Epstein-Barr virus RNAs  
AUTHORS Nucleic Acids Res. 15 (14), 5887 (1987)  
87289053

JOURNAL 16 (bases 1 to 172281)  
MEDLINE Laux, G., Perricaudet, M. and Farrell, P.J.  
REFERENCE A spliced Epstein-Barr virus gene expressed in immortalized  
AUTHORS lymphocytes is created by circularization of the linear viral  
TITLE genome  
EMBO J. 7 (3), 769-774 (1988)  
88283646

JOURNAL 17 (bases 1 to 172281)  
MEDLINE Hatfull, G.F., Barrell, B.G., Quin, J. and McGeoch, D.  
REFERENCE Unpublished  
AUTHORS unpublished  
TITLE unpublished

JOURNAL 18 (bases 1 to 172281)  
MEDLINE Farrell, P.J. and Barrell, B.G.  
REFERENCE Direct Submission

JOURNAL  
REFERENCE 19 (bases 1 to 172281)  
AUTHORS Farrell,P.J.  
TITLE Direct Submission  
JOURNAL Submitted (18-MAR-1988) Farrell P., Ludwig Institute for Cancer Research, St. Mary's Hospital Medical School, Norfolk Place London W2 1PG  
COMMENT CDS

Listed under this feature are all known protein coding regions as well as all the major open reading frames in the sequence. In particular the term major is taken as the longest frame in a particular region taking into account the adjacent longest frames and likely transcription signals. Note that on this basis some long overlapping frames have been excluded and on the other hand some small frames have been included which might represent exons or genes because they occur in a logical combination with other features or because of some other experimental data. The reading frames are named according to the Bam HI fragment in which they start. eg BALF3 is the third leftward frame starting in Bam HI fragment A. BORE1 is the first rightward frame in Bam HI fragment O. If there is an obvious TATA sequence followed by an in frame Met codon that satisfies the rules of Kozak [12] in that there is a purine at -3 and/or a G at +4 then the reading frame is numbered from the A of the ATG to the base preceding the termination codon. If there is no obvious initiation codon or there is a substantial reading frame in phase before the ATG then the reading frame is numbered from the first base of the first codon.

## SITES of POLYA signals

This feature lists all occurrences of the sequence AATTA which is found normally approximately 20 bases upstream of the mRNA processing/polyA addition site. The rarely used homolog AATTA is only listed when it is found in a position close to the end of a major reading frame.

## SITES of DONOR and ACCEPT sequences

This is not a comprehensive listing of all such sequences and only the positions of a few have been noted because they occur in potentially interesting positions. The number quoted in the table is the position of the terminal base in the intron in each case.

## Restriction enzyme SITES.

Only the positions of the sites Bam HI (BAM) are listed.

## RPT

This feature is used to define repetitive sequences.

## SITE DEL

This feature defines deletions in B95-8 with respect to other strains such as RAJ1 and also to deletions in other strains such as PHR1 and DAUDI with respect to B95-8.

## SITE HPN

Denotes sequences with twofold symmetry ie could form hairpin loops. This is not a comprehensive list - only a few occurrences noted.

## ORGRPT

Denotes the region that encompasses an origin of replication (ori P). [13].

## NUMBERING

The DNA sequence of B95-8 EBV has been revised [19]. The original (Baer et al, 1984) base 359 has been deleted so the new sequence around that position reads TCAGTCCTT. To avoid renumbering the entire sequence, position 1 has been moved 1 base to the left of the EcoRI site separating EcoRI Dhet from EcoRI I (ie the first A of AGAATTC).

Location/Qualifiers

1. 172281

/organism="Human herpesvirus 4"

/strain="B95-8"

/db\_xref="taxon:10376"

58..272

/note="exon 2 terminal protein RNA"

360..458

/note="exon 3 terminal protein RNA"

complement(535)

/note="polya signal: AATTA"

540..788

/note="exon 4 terminal protein RNA"

871..951

/note="exon 5 terminal protein RNA"

1026..1196

/note="exon 6 terminal protein RNA"

complement(1192)

/note="TATA: TATTAAT"

1280..1495

/note="exon 7 terminal protein RNA"

complement(1383)

/note="TATA: CATTAATA"

1574..1682

/note="exon 8 terminal protein RNA"

1676

/note="TATA: TATAAAG"

1691

/note="TATA: TATTAATA BN-R1 late promoter before BNR1, gives 4.1kb late RNA. Probably encodes non glycosylated 140kd protein in membrane antigen. Also two latent RNAs spliced underneath this RNA, lengths 1.8 and 2.0kb (Hudson et al, 1985). The longer one encodes terminal protein."

1736..5692

/note="BNR1 reading frame, 5' NKT/S"

/codon\_start=1

/db\_xref="PID:959075"

/translation="MERGEMOMPVARYGPPIMVIRFGDGEANTIOBERLYELSD

PRSAIGLDGPEPLAENLILVALRGTYNDPRPOQREARALALVIGILGNEOSEHIGT  
ESALERSGNTVYATIGPDWARSSTWSAETIOQLRLIGATYVLRVEGRQFGEVHS  
RPSERFOALNHLVATFDNALRKDYDSQVAAAGFORALVAGPETAADRPRKLNVEF  
GGRAGAGROIALDELKIVSALRDYVSGHVLQPEETLDTWKLSRDTRTASLHGFIH  
AAGTIOANCPOLFMROHGLPEFVNALSIASLGWYOTATPGADARAARROAOOT  
RAABCHAKSGVAVAGFYRTIATLKGEGLOPTMENGELGAKHOADHTVYDGH  
YLIMGPFOPMSGLTAPPCPTAASSWAOAAVOTALEFSALTYAPCTISGARPPGSA  
VIEHLSIVPKGLLFLSHLPDVKDGLGEMKPARATGMOQFVSSYFLNACSNV  
FIVRORGEKINRIVYALGRACDAGCOHAYGVSTVPLGINTFNDLASPVSTAEH  
MDPSEFFVEEPIOEAGSSVPLDVEDSMOISPSYELPMLSLSCSLISHPTV  
GSKHELVRTDRVSGRVAQOQVGLDPLADYAVAVAHQVWTRGAGAPLRYRTD  
RMTEKILYSAKPGGKGVKSGVYITIGEGGKYSLDRGTRILAMAEILNACAPIL  
DPEDVYLITLHLIDPRADNSAVMEAMTASDARGLCYALITGSGASCPTGSSASNF  
MTVAVASVAPGERSGPLTIVLOKTSGLLAVRCGDKIOGSLFPOLEVDVATTPRA  
PEALSLKNEFRAVOQLVKSGLVLSGHDISDGLVYCLVEAMLAGRGVITTMVADY  
LPMPEAHEBGLVEVEERSYGEVLQTLRINSMMYPAVGRVGEQPGOMEQVGPETV  
LROSILRLIGTMSFASBOYECPLPRINSMMYPAVGRVGEQPGOMEQVGPETV  
VPEPDRCOYAVICAPGTGRHESLILAAFNAGLCGRVPRERAVDTPDYKGLAIG  
GVHARDASALAGRATYALINRPPALDAITKILNRPDTSVALGELGVYALGAVG  
STNPPAPGVENVVORSPLILAPNASGMESRWLNTSIPATISSVLRGLRGVGLCW  
VOGSCGLQIOTNIGMEYVLONAHQIACHFSRSGCTDAMRFRAMNPNRTGNGTAGCS  
RDGRHALLCDPSTICTDFWQMEHIIPAFGHPTGCSWPWTLMFOAHILMSIRHGRPSE"

```

misc_feature complement(1795)
/note="polya signal: AATAAA"
misc_feature 3955
/note="BAM: Bam HI NheI/h"
misc_feature 3994
/note="BAM: Bam HI h/C"
5408..5856
/note="exon 9 terminal protein RNA"
misc_feature 5841
/note="polya signal: AATAAA, end of 4.1kb late RNA and TP
latent RNA."
misc_feature 5863
/note="alternative end to TP CDNA5"
promoter 6097
/note="TATA: TATAAGA"
misc_feature 6629..6795
/note="pol III RNA EBER 1"
complement(6823)
promoter /note="TATA: CATAAAT"
6956..7128
/note="pol III RNA EBER 2"
7315..9312
/note="origin of replication, ori P (Yates et al, 1984,
1985)"
7421..8042
/note="21x30bp repeats, binding sites for EBNA-1 (site I,
Rawlins et al, 1985). Tandem repeat part of orip (Reisman
et al, 1985). Also functions as a cell type specific
enhancer (Reisman et al, 1985; Lupton and Levine, 1985)"
7738
promoter /note="TATA: TATAAAT"
7888
promoter /note="TATA: TATAAAT"
8573
/note="TATA: CATAAAT"
misc_feature /note="TATA: CATAAAT"
complement(8680)
misc_feature /note="polya signal: AATAAA"
complement(8755)
misc_feature /note="polya signal: AATAAA"
8962
/note="polya signal: AATAAA"
9021..9133
/note="HPN: dyad symmetry, site II for EBNA-1 binding
(Rawlins et al, 1985). Dyad symmetry part of orip (Reisman
et al, 1985)"
misc_feature complement(9398)
promoter /note="TATA: TATAAAT"
9631
/note="TATA: TATAAAT BC-RI late promoter before BCRF1"
9675..10187
CDS /note="BCRF1 reading frame"
/note="BCRF1
/codon_start=1
/db_xref="PID:g59076"
/db_xref="SWISS-PROT:P03180"
/translaton="MERLVVTLQCLVLYLTAPECGGTDCDNFPQMLRIDAFSRV
KTFQTKDEVNLLKESLDFKGLQSLSEIDIOFLVEVMPQANODPEAKDHY
NSLGENKTLRLRLRCHRFLEPCENKSAVQIKNAENKLEKGIYKMSBDFIINY
IEAVMTIKAR"
promoter 10076
/note="TATA: GATAAAA"
misc_feature complement(10148)
/note="polya signal: AATAAA"
10173
/note="polya signal: AATAAA"
misc_feature /note="polya signal: AATAAA"
10257
/note="polya signal: AATAAA, end of 0.8kb late RNA from
BCRI and end of 1.6 kb late RNA, start unknown"
misc_feature complement(10277)
promoter /note="polya signal: AATAAA"
complement(10975)
/note="TATA: CATAAAT"
11305
/note="TATA: TACAATA: BCR2 promoter for highly spliced

```

```

mRNA EBNA latent RNAs."
11336..11480
/note="exon C1 of Bodescot et al (1986) RNAs"
promoter 11524
/note="TATA: TATAATT"
complement(11587)
misc_feature /note="polya signal: AATAAA"
complement(11606)
promoter /note="TATA: CATAAAT"
11626..11657
/note="exon C2 of Bodescot et al (1986) RNAs"
promoter 11796
/note="TATA: TATAAGT"
complement(11799)
promoter /note="TATA: TATAAAA"
12001..15072
/note="3072 repeat 1"
12541..13689
/note="BCRF2 3072 repeat, reading frame 1; Protein
sequence is in conflict with the conceptual translation"
CDS /db_xref="PID:e25077"
/db_xref="PID:g1334836"
/db_xref="SPTRMBL:004397"
/translaton="VWEABGRPRPGEVEGDRLPCLQSPEDLPRLSPGSGRSAPOTD
PRVSRGPPASSGAAGSPPOAPOTRVASRADRRARMLGASRRGFCPSLCSBEG
TSGTPERLGPASRRPPLGSLSPVSKRECLRGATLGAQAPESGGHLLVPRVPCQ
PECPRGGRQRPVPRPGLQSGPCPEGTGLVSPPLQARASPSRSGASLGPQVQ
HRDPSGDPPTGSLCPAPLQPSLHPRQLASVSPGPGPEGRPGRAVFLPML
LPASHPSPLSPRHVHQAQRDPGGSVPPAAAGSLPGKGSFSPSLRSLCT
VCKVQPTPVHGSRAQRPPLPTVDRSVHPRPVSTVPSRQDFM"
13215
misc_feature /note="BAM: BamHI C/W"
14352
/note="**TATA: TATAAG BMRI one of the promoters for highly
spliced EBNA and LP RNAs (Sample et al, 1986; Speck et al,
1986)"
14384..14410
/note="**exon W0 of EBNA/LP RNAs"
14554..14619
/note="**exon W1 (also W66) part of leader protein (LP)
gene. LP is also called EBNA-5 (Dillner et al, 1986) and
EBNA-4 (Rowe et al, 1987)."
14559..14619
/note="**exon W1' (also W61) of EBNA/LP RNAs forms
initiator met when fused to exon W0 or exon C2."
14701..14832
/note="**exon W2 (also W132) part of LP gene"
15073..18144
repeat_region /note="3072 repeat 2"
15613..16761
CDS /note="BMRF1 reading frame 2; Protein sequence is in
conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25078"
/db_xref="PID:g1334837"
/db_xref="SPTRMBL:004397"
/translaton="VWEABGRPRPGEVEGDRLPCLQSPEDLPRLSPGSGRSAPOTD
PRVSRGPPASSGAAGSPPOAPOTRVASRADRRARMLGASRRGFCPSLCSBEG
TSGTPERLGPASRRPPLGSLSPVSKRECLRGATLGAQAPESGGHLLVPRVPCQ
PECPRGGRQRPVPRPGLQSGPCPEGTGLVSPPLQARASPSRSGASLGPQVQ
HRDPSGDPPTGSLCPAPLQPSLHPRQLASVSPGPGPEGRPGRAVFLPML
LPASHPSPLSPRHVHQAQRDPGGSVPPAAAGSLPGKGSFSPSLRSLCT
VCKVQPTPVHGSRAQRPPLPTVDRSVHPRPVSTVPSRQDFM"
16287
/note="BAM: BamHI W/W"
17424
promoter /note="TATA: TATAAG"
17626..17691
mRNA /note="Exon W1"
17773..17904
mRNA /note="Exon W2"
18145..21216
repeat_region

```

```
CDS
    /note="3072 repeat 3"
    /note="18685..19833
    /note="BWRf1 reading frame 3: Protein sequence is in
    conflict with the conceptual translation"
    /codon_start=1
    /db_xref="pid:e25079"
    /db_xref="pid:g1334838"
    /db_xref="SPTREMBL:Q04397"
    /translation="VWEAGRRPRGVEGDRPGLCMQSPGDFLRLPSGGRSPSPAPOTD
    PVSROGPASSGAGSPPOAPOTRVASRADRPRAWRLLGASRGWFCPSLCPSEEPG
    TSGTEPLGPARRRPGLRSLSPVKPKKCLGATLGAQAPESRGGHRLRVPRVPGO
    PGPRQGPGRQVRPRPRPGLQSPGCPREGTIGVSPRLQARASPSRRKASIGPOVOP
    HNDPSGDPPTGSLCPPAPLQPSLHPRQLLASGPPGQPGPRQPGKVAFLPLWPL
    LPASHPSPLSLPRHVHQAGRDPCGPGVSVPPAAQSLPPGKASFSPPSLRLSLCT
    VCKVQPPPTPVHGSRAQRPRLPTVDRPSVHPGHPRPVPSTPVPSRGDFW"
    19359
misc_feature
    /note="BAM: BamH1 W/W"
    20496
promoter
    /note="TATA: TATAAG"
    20698..20763
mrna
    /note="Exon W1"
    20845..20976
    /note="Exon W2"
    21217..24288
repeat_region
    /note="3072 repeat 4"
    21757..22905
CDS
    /note="BWRf1 reading frame 4: Protein sequence is in
    conflict with the conceptual translation"
    /codon_start=1
    /db_xref="pid:e25080"
    /db_xref="pid:g1334839"
    /db_xref="SPTREMBL:Q04397"
    /translation="VWEAGRRPRGVEGDRPGLCMQSPGDFLRLPSGGRSPSPAPOTD
    PVSROGPASSGAGSPPOAPOTRVASRADRPRAWRLLGASRGWFCPSLCPSEEPG
    TSGTEPLGPARRRPGLRSLSPVKPKKCLGATLGAQAPESRGGHRLRVPRVPGO
    PGPRQGPGRQVRPRPRPGLQSPGCPREGTIGVSPRLQARASPSRRKASIGPOVOP
    HNDPSGDPPTGSLCPPAPLQPSLHPRQLLASGPPGQPGPRQPGKVAFLPLWPL
    LPASHPSPLSLPRHVHQAGRDPCGPGVSVPPAAQSLPPGKASFSPPSLRLSLCT
    VCKVQPPPTPVHGSRAQRPRLPTVDRPSVHPGHPRPVPSTPVPSRGDFW"
    22431
misc_feature
    /note="BAM: BamH1 W/W"
    23568
promoter
    /note="TATA: TATAAG"
    23771..23835
mrna
    /note="Exon W1"
    23917..24048
    /note="Exon W2"
    24289..27360
repeat_region
    /note="3072 repeat 5"
    24829..25977
CDS
    /note="BWRf1 reading frame 5: Protein sequence is in
    conflict with the conceptual translation"
    /codon_start=1
    /db_xref="pid:e25081"
    /db_xref="pid:g1334840"
    /db_xref="SPTREMBL:Q04397"
    /translation="VWEAGRRPRGVEGDRPGLCMQSPGDFLRLPSGGRSPSPAPOTD
    PVSROGPASSGAGSPPOAPOTRVASRADRPRAWRLLGASRGWFCPSLCPSEEPG
    TSGTEPLGPARRRPGLRSLSPVKPKKCLGATLGAQAPESRGGHRLRVPRVPGO
    PGPRQGPGRQVRPRPRPGLQSPGCPREGTIGVSPRLQARASPSRRKASIGPOVOP
    HNDPSGDPPTGSLCPPAPLQPSLHPRQLLASGPPGQPGPRQPGKVAFLPLWPL
    LPASHPSPLSLPRHVHQAGRDPCGPGVSVPPAAQSLPPGKASFSPPSLRLSLCT
    VCKVQPPPTPVHGSRAQRPRLPTVDRPSVHPGHPRPVPSTPVPSRGDFW"
    25503
misc_feature
    /note="BAM: BamH1 W/W"
    26640
promoter
    /note="TATA: TATAAG"
    26842..26907
mrna
    /note="Exon W1"
    26989..27120
    /note="Exon W2"
    27361..30432
repeat_region
    /note="3072 repeat 6"
    27901..29049
CDS
    /note="BWRf1 reading frame 6: Protein sequence is in
    conflict with the conceptual translation"
    /codon_start=1
    /db_xref="pid:e25066"
    /db_xref="pid:g1334841"
    /db_xref="SPTREMBL:Q04397"
    /translation="VWEAGRRPRGVEGDRPGLCMQSPGDFLRLPSGGRSPSPAPOTD
    PVSROGPASSGAGSPPOAPOTRVASRADRPRAWRLLGASRGWFCPSLCPSEEPG
    TSGTEPLGPARRRPGLRSLSPVKPKKCLGATLGAQAPESRGGHRLRVPRVPGO
    PGPRQGPGRQVRPRPRPGLQSPGCPREGTIGVSPRLQARASPSRRKASIGPOVOP
    HNDPSGDPPTGSLCPPAPLQPSLHPRQLLASGPPGQPGPRQPGKVAFLPLWPL
    LPASHPSPLSLPRHVHQAGRDPCGPGVSVPPAAQSLPPGKASFSPPSLRLSLCT
    VCKVQPPPTPVHGSRAQRPRLPTVDRPSVHPGHPRPVPSTPVPSRGDFW"
    28575
misc_feature
    /note="BAM: BamH1 W/W"
    29712
promoter
    /note="TATA: TATAAG"
    29914..29979
mrna
    /note="Exon W1"
    30061..30192
    /note="Exon W2"
    30433..33504
repeat_region
    /note="3072 repeat 7"
    30973..32121
CDS
    /note="BWRf1 reading frame 7: Protein sequence is in
    conflict with the conceptual translation"
    /codon_start=1
    /db_xref="pid:e25067"
    /db_xref="pid:g1334842"
    /db_xref="SPTREMBL:Q04397"
    /translation="VWEAGRRPRGVEGDRPGLCMQSPGDFLRLPSGGRSPSPAPOTD
    PVSROGPASSGAGSPPOAPOTRVASRADRPRAWRLLGASRGWFCPSLCPSEEPG
    TSGTEPLGPARRRPGLRSLSPVKPKKCLGATLGAQAPESRGGHRLRVPRVPGO
    PGPRQGPGRQVRPRPRPGLQSPGCPREGTIGVSPRLQARASPSRRKASIGPOVOP
    HNDPSGDPPTGSLCPPAPLQPSLHPRQLLASGPPGQPGPRQPGKVAFLPLWPL
    LPASHPSPLSLPRHVHQAGRDPCGPGVSVPPAAQSLPPGKASFSPPSLRLSLCT
    VCKVQPPPTPVHGSRAQRPRLPTVDRPSVHPGHPRPVPSTPVPSRGDFW"
    31647
misc_feature
    /note="BAM: BamH1 W/W"
    32784
promoter
    /note="TATA: TATAAG"
    32986..33051
mrna
    /note="Exon W1"
    33133..33264
    /note="Exon W2"
    33505..36576
repeat_region
    /note="3072 repeat 8"
    34045..35193
CDS
    /note="BWRf1 reading frame 8: Protein sequence is in
    conflict with the conceptual translation"
    /codon_start=1
    /db_xref="pid:e25068"
    /db_xref="pid:g1334843"
    /db_xref="SPTREMBL:Q04397"
    /translation="VWEAGRRPRGVEGDRPGLCMQSPGDFLRLPSGGRSPSPAPOTD
    PVSROGPASSGAGSPPOAPOTRVASRADRPRAWRLLGASRGWFCPSLCPSEEPG
    TSGTEPLGPARRRPGLRSLSPVKPKKCLGATLGAQAPESRGGHRLRVPRVPGO
    PGPRQGPGRQVRPRPRPGLQSPGCPREGTIGVSPRLQARASPSRRKASIGPOVOP
    HNDPSGDPPTGSLCPPAPLQPSLHPRQLLASGPPGQPGPRQPGKVAFLPLWPL
    LPASHPSPLSLPRHVHQAGRDPCGPGVSVPPAAQSLPPGKASFSPPSLRLSLCT
    VCKVQPPPTPVHGSRAQRPRLPTVDRPSVHPGHPRPVPSTPVPSRGDFW"
    34719
misc_feature
    /note="BAM: BamH1 W/W"
    35836
promoter
    /note="TATA: TATAAG"
    36058..36123
mrna
    /note="Exon W1"
    36205..36336
    /note="Exon W2"
    36577..39648
repeat_region
    /note="3072 repeat 9"
    37117..38265
CDS
```

/note="BMRFL reading frame 9; Protein sequence is in  
conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25069"  
/db\_xref="PID:g1334844"  
/db\_xref="SPTRMBL:004397"  
/translation="VWEAGRPREGVEGDRPGICWQSPDDPLRPSGPGSPSPAPQTD  
PRVSRQGPASSGAGSPPOAQTFRVSASRADRPAMWLLGASRGWFCPSLCPSEEG  
TSGTPEPLGASRRPGLRSPLSVKPKCELRGATLGAQAPESRGQCHLVRPVVQ  
PEGROGRRPQRPVPRPGLSPGCPCEGLGVSPPLQARASPSRSGASLGVOVP  
HRDPSGDPPTGSLCPAPLQPSLHRPOLLASPGPPGPEGROGRRVAFPLPML  
LPASHPSPLSLPRHVVHQAQRDQGVSYPPAAQSLPPGKAGSESPSLRSLTCT  
VCKVQPTPVHGSRAQRPPLPYDRPSVHGHPRPVSTVPBKGDFM"  
37791  
misc\_feature  
/note="BAM: BamH1 W/M"  
38928  
/note="TATA: TATAAG"  
39130..39195  
/note="Exon W1"  
39277..39408  
/note="Exon W2"  
39649..42720  
/note="3072 repeat 10"  
40189..44337  
/note="BMRFL reading frame 10; Protein sequence is in  
conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25070"  
/db\_xref="PID:g1334845"  
/db\_xref="SPTRMBL:004397"  
/translation="VWEAGRPREGVEGDRPGICWQSPDDPLRPSGPGSPSPAPQTD  
PRVSRQGPASSGAGSPPOAQTFRVSASRADRPAMWLLGASRGWFCPSLCPSEEG  
TSGTPEPLGASRRPGLRSPLSVKPKCELRGATLGAQAPESRGQCHLVRPVVQ  
PEGROGRRPQRPVPRPGLSPGCPCEGLGVSPPLQARASPSRSGASLGVOVP  
HRDPSGDPPTGSLCPAPLQPSLHRPOLLASPGPPGPEGROGRRVAFPLPML  
LPASHPSPLSLPRHVVHQAQRDQGVSYPPAAQSLPPGKAGSESPSLRSLTCT  
VCKVQPTPVHGSRAQRPPLPYDRPSVHGHPRPVSTVPBKGDFM"  
40863  
misc\_feature  
/note="BAM: BamH1 W/M"  
42000  
/note="TATA: TATAAG"  
42202..42267  
/note="Exon W1"  
42349..42480  
/note="Exon W2"  
42721..45792  
/note="3072 repeat 11"  
43261..44409  
/note="BMRFL reading frame 11; Protein sequence is in  
conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25071"  
/db\_xref="PID:g1334846"  
/db\_xref="SPTRMBL:004397"  
/translation="VWEAGRPREGVEGDRPGICWQSPDDPLRPSGPGSPSPAPQTD  
PRVSRQGPASSGAGSPPOAQTFRVSASRADRPAMWLLGASRGWFCPSLCPSEEG  
TSGTPEPLGASRRPGLRSPLSVKPKCELRGATLGAQAPESRGQCHLVRPVVQ  
PEGROGRRPQRPVPRPGLSPGCPCEGLGVSPPLQARASPSRSGASLGVOVP  
HRDPSGDPPTGSLCPAPLQPSLHRPOLLASPGPPGPEGROGRRVAFPLPML  
LPASHPSPLSLPRHVVHQAQRDQGVSYPPAAQSLPPGKAGSESPSLRSLTCT  
VCKVQPTPVHGSRAQRPPLPYDRPSVHGHPRPVSTVPBKGDFM"  
43935  
misc\_feature  
/note="BAM: BamH1 W/M"  
45072  
/note="TATA: TATAAG"  
45274..45339  
/note="Exon W1"  
45415..52824  
/note="DEL: DAUDI deletion (Jones et al, 1984)"  
45421..45552  
/note="Exon W2"  
45644..52450  
/note="DEL: P3H1 deletion (Jeang and Hayward, 1983)"

repeat\_region  
/note="3072 repeat 12"  
46333..47481  
/note="BMRFL reading frame 12; Protein sequence is in  
conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID:e25072"  
/db\_xref="PID:g1334847"  
/db\_xref="SPTRMBL:004397"  
/translation="VWEAGRPREGVEGDRPGICWQSPDDPLRPSGPGSPSPAPQTD  
PRVSRQGPASSGAGSPPOAQTFRVSASRADRPAMWLLGASRGWFCPSLCPSEEG  
TSGTPEPLGASRRPGLRSPLSVKPKCELRGATLGAQAPESRGQCHLVRPVVQ  
PEGROGRRPQRPVPRPGLSPGCPCEGLGVSPPLQARASPSRSGASLGVOVP  
HRDPSGDPPTGSLCPAPLQPSLHRPOLLASPGPPGPEGROGRRVAFPLPML  
LPASHPSPLSLPRHVVHQAQRDQGVSYPPAAQSLPPGKAGSESPSLRSLTCT  
VCKVQPTPVHGSRAQRPPLPYDRPSVHGHPRPVSTVPBKGDFM"  
47007  
misc\_feature  
/note="BAM: BamH1 W/Y"  
47761..47793  
/note="Exon Y1 Bodescot et al, 1984"  
47831  
/note="TATA: TATAAGT"  
47878..47999  
/note="Exon Y2 Bodescot et al, 1984 and EBNA-1 (Speck and  
Strominger, 1985), last common exon"  
complement(48023)  
/note="polya signal: AATAAA"  
48386..48444  
/note="exon Bodescot et al, 1984"  
48386..50032  
/note="Coding exon for EBNA-2 (Sample et al, 1986)"  
/codon\_start=1  
/db\_xref="PID:e25073"  
48504..49667  
/note="BMRFL encodes EBNA-2 (Dambaugh et al, 1984);  
Dillner et al, 1984)"  
/codon\_start=1  
/db\_xref="PID:e25074"  
/db\_xref="PID:g1632787"  
/db\_xref="SWISS-PROT:P12978"  
/translation="MPTFYALHGGOTYLIYDTSIGNPSLVISNRYOGLSDTP  
LIPLITFVGENGVPPPLPPLPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPPP  
RDAMQEPSPDLDDPLGYDVGHGLASAMKIMANVIVQSGKDGILLPOQYAP  
QARLVQPHVPLRPATPILISPLSQPLRPPLQILMPPRPPTPLPPLATILVPRPT  
RPTTLRPPLTLVLRTELOQPTSPRMHLVYLAVDQSMHPLTHQSPNDPDSPEP  
RSPTFYVNIIPMPLPSPOLPPRAAPQRPQVINDOGLHLHLPGWMPIDCPQPS  
KIOGQSGOSRGGRGGRGKGRKGRKRDQKRGKGRWREPNSTSSMELSVGLH  
OGGAGDSPTPGPSNAAPVCNSHATAPNVSPKIHPESSHNSPAPILFPDDWTPSID  
PADLDSWDYIEFTTESPSSDDEYVGGPKRPRPSIO"  
48678..48800  
/note="14 x CCCCACCA repeats"  
48848  
/note="BAM: BamH1 Y/H"  
49350  
/note="TATA: TATAACA"  
complement(49353)  
/note="TATA: TATAAAA"  
49525..49578  
/note="9 x GGGCCA repeats"  
49852..50032  
/note="exon (Bodescot et al 1984)"  
50003  
/note="polya signal: AATAAA, end of Bodescot T1 RNA and  
EBNA-2 RNA (3.0kb latent RNA in IB4 cells)"  
complement(50156)  
/note="TATA: TATAAGT"  
/note="polya signal: AATAAA, end of 2.5kb early RNA from  
52817"  
complement(50578..52577)  
/note="BMRFL early reading frame"  
50578..52115  
/note="12 x 125bp repeats"



misc\_feature 52654..53697  
/note="region homologous to Eco RI C of Raji"  
promoter complement(52817)  
/note="TATA: GATAAA promoter for 2.5kb early RNA  
containing BHLF1 (Jeang and Hayward, 1983; Fresse et  
al,1983)"  
promoter 53759  
/note="TATA: TATTAA likely promoter for class III and IV  
early RNAs encoding BHRF1 (Pearson et al, 1987)"  
misc\_feature 53895  
/note="DONOR: CCGGACT donor for splice to 54335 in class  
IV early RNAs encoding BHRF1 (Pearson et al, 1987)"  
misc\_feature 54335  
/note="ACCEPT: TTTCTAG acceptor from 48444 in class I,  
47999 in class II, and 53895 in class IV early RNAs  
encoding BHRF1 (Pearson et al, 1987)"  
misc\_feature 54376..54948  
/note="BHRF1 reading frame, limited homology to bcl-2 gene.  
Early gene in B95-8 cells and part of restricted EA  
complex."  
promoter 54591  
/note="TATA: TATAACA"  
promoter complement(54594)  
/note="TATA: TATTAAT"  
misc\_feature 54853  
/note="BAM: BamH1 H/F"  
misc\_feature complement(54929)  
/note="polya signal: AATAAA"  
promoter complement(54977)  
/note="TATA: TATAAAG"  
misc\_feature 5518  
/note="polya signal: AATAAA, 3' end of 2.5kb, 1.9kb, 1.7kb  
and 0.6kb early RNAs"  
misc\_feature complement(55982..56935)  
/note="BFLF2 reading frame, 4 NXT/S, homologous to RF 27  
in VZV and HRF2 in CMV"  
misc\_feature complement(55990)  
/note="polya signal: AATAAA, 3' end of 2.3kb and 1.1kb  
early RNAs from 58568 and 57081"  
promoter complement(56132)  
/note="TATA: TATAAAG"  
misc\_feature complement(56948..58525)  
/note="BFLF1 reading frame, 2 NXT/S homologous to RF 26 in  
VZV and HRF1 in CMV"  
/db\_xref="PID:g1334849"  
/db\_xref="SWISS-PROT:P03184"  
/translation="MAKVTISANEPNULGKRLSCPLTRSGVTEVAQIAGRTPKMED  
FVPMVDNLSKSOEAVGLMAHSYLPANAEGLAYPLVHTYESLSPASTCRCDLID  
TLVNSDAPVAFEDYALCYCLNAPRAVISLITGMDFLHILIKYFPAGGLISF  
MPSRLAIDILHFIYRCFLPYSSSDMINANMIGYKLEFLKSLTGOSPANCFKS  
MPRTPTFLTPGPRCKSDOPDVGNGRLTALCCHPTRNVOHPHTRAKGGL  
SPEITTKADYCGLLIGTWOCTDLGGPRGHAIISMEYSGDELAELALATTPREAGDH  
SOGPCLPAGHGHKKNASRTICPLCSLADHPADAKDTDRKSLIDSGNNIKILD  
RIYFLIKTONTLLDVPCLRLANLQMCPTPODFKHLCFCDLCAINHSITNPVLFQI  
YPSQAFKALALAGCNLEQGVCDSLITLVYIFKSTQVAVKRTIIVDYTKELDVVLR  
IHLDLVQSYQTSOVY"  
promoter complement(57081)  
/note="TATA: TATTAA before BFLF2; BFL2 promoter gives  
1.1kb early RNA"  
promoter complement(58088)  
/note="TATA: GATAAA"  
promoter complement(58568)  
/note="TATA: TATTAA before BFLF1, BFL1 promoter gives  
2.3kb early RNA"  
promoter 58832  
/note="TATA: TATAAAA before BFRF1"  
CDS 58891..59901  
/note="BFRF1 early reading frame, 1 NXT/S, homologous to  
HFLF4 in CMV"  
/codon\_start=1  
/db\_xref="PID:g1334850"  
/db\_xref="SWISS-PROT:P03185"

CDS 59808..61583  
/note="BFRF2 early reading frame, homologous to HFLF5 in  
CMV"  
/codon\_start=1  
/db\_xref="PID:e25058"  
/db\_xref="PID:g1632788"  
/db\_xref="SWISS-PROT:P14347"  
/translation="MALFLARHTLSTGTCGCHGRGAPDVEYDLTLOALGERFSRL  
LDLGLDLSVYEMREPFYWGWRPASPAVAATPGLSPHSAYWISPEVRRGLY  
RNAROTSRSLSLALHLPGLSLSLSVYTGCGGAGVAFCAFLAFLVLYRSVPFAPS  
DRIANICDSRPFCEPENTRAVARGYRGVLVFLAVERSYDPLLRONSRAVERCFA  
IKNYVGLDSQSCVITVPSRMAQSHASELDEPRIDRYVPATAPSFVDHASALLAS  
LKKASDTEPCGNIDMMWYRLVNSCLRSACLPIPAVSEGGRTGGGVELVAGAG  
GPTLSRDVFAIVSRNVLSCILNVPAPGAPRAYKCFRSHASRPVSGPDYAFACDC  
GYCLNFKQKOTGVGRSLNSFRPTLQFYPROKENVHITCHASGRVYSCNGSAVGCOR  
LAEPARSQWRPRITAVLPNNAIETLDKSRLLDHIIPGLPDRTCMRPVILRGVTR  
KOLLVLTLTLEARAVCSICQROAPEDADDEPHLFSSCLEVELPEERCAGRLYQTR  
YGPAAQAHPPGEGAGGFSRSPAS"  
promoter complement(61062)  
/note="TATA: GATAAA"  
promoter 61344  
/note="TATA: TATTAA before BFRF3"  
CDS 61507..62037  
/note="BFRF3 early reading frame"  
/codon\_start=1  
/db\_xref="PID:e25059"  
/db\_xref="PID:g1632789"  
/db\_xref="PID:g1632789"  
/translation="MARLKEPPTLQRLADFPDPSLPLKFOELNONNLPNDVFRBAQ  
RSYLVLTSGFCYEYVORTFVGPORRAIDKROASVAGCAHAHLGSSATPVQQA  
QAAASAGTGALASASPTVAQSATPSVSSISLNAATSGATMAASAAAVDTGSGG  
GGPHDTAPGAKRKO"  
misc\_feature complement(62068)  
/note="polya signal: AATAAA"  
misc\_feature 62069  
/note="polya signal: AATAAA, 3' end of 10, 6.5, 3.7, 3.4,  
3.1, 2.5 and 0.8kb early RNAs"  
misc\_feature complement(62078..71527)  
/note="BPLF1 reading frame, 1 NXT/S, analogous to VZV  
RF2"  
/codon\_start=1  
/db\_xref="PID:g1334853"  
/db\_xref="SWISS-PROT:P03186"  
/translation="NSNGMGOQQRTRGRGVPVGRITMDVYNAFGGSGSALRICTA  
SCQNAKCRGRPGICTCVSNVLIYKSLAGRPVLSRELDVLDGALDLMKOS  
GILKGHEMOQLDVPSSVLRGGGRVHIFRSALFGLVLEPAQIANSAAVOISAEVLH  
GSYNGAAGYILYICDITYAGAIILIEDGSFLEFPHCOKDAAPETPAHVRYSIAEHL  
QYVAGAGAYTCHLYLEAEFTEDEPRTFMLEHLYVDFEYEGSGFDLVCBELVS  
DGAAGTGGDSSPYMLPFEBRRIILYNNLRPLRSFTSDSFPAAARYSPAKTNPSS  
PASAASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPASAPAS  
GYPABSTPPASGGAAPOTPKRKGGLKQSPKRPISGRPLUSTTDEDDLPPTH  
VPHRPPSARLPAPVPIPHOSPPASPPHPAPVSTIAPSVPSRPLQPIPIPO  
AAPSNNKIPITPPSPSPTAAATTTTLSPPOQOPPOSAAAPSLPPOQOPTPSA  
APASPLPPOQOPPSARAPSLPPOQOPPLPPTAPAPPAQOLPPSATLLEKKNP  
PADRAGTEISPSPPRGOOPSGEDASGGLVRLDLEPPLMSDEBEASDLAS  
DIPTEDDEMEFDEVSNSLESSSNPSTIYLDYRISOYTOTPIETEPEDAFVLE  
SNIRIARAGHTIOEQAIVTPASRREVPEDADLSMIYLLVTVLEGLIRSRORSVL  
NLLEFLKDSNGHLQVPTLDEOLLITSELINQNLANNLENGKAGGFHHLAKLEAC  
LPSIARDAVNDVADAGAKMLEIPLAESDGDGKDLDAARRLLDLSGDOEGEGCG  
EPEDNSIYGPHYDVPLVLDSEMKRLSLAEARAVARQAGVEEDVREFALITTA  
IEYGAPPASVSPPEVNVAVRSKNALHRLRCGTADIRDVASASDYLSYLEDPSLT  
VMDFDLLHRLHTCOIILASPLINTRYISIMDYVELLYGTALSDMSGIMPIDERV  
EEDDPSIARLPETETAKOKELFTTRREMKRRTLITDIEMGLAGVASAPGATIS  
PASPSATPANHNDPEATPLADTALTYIVLEKTYIANAGSIYGAAKNPYIRLRTIO  
QVIRSKRYLMNLKISITYTIDNYIASFEESIDHLTRAGSVLDVDOEDIDRLDPMV  
SEALHTFEMGNRLTEPARVALONFATSTLKETAAVANLLPGLLAVDATITGQAP







```

misc_feature
87650
/note="BAM: Bam H1 S/L"
88507
/note="TATA: TATAAT BL-R1 late promoter before BLRF1,
gives 1.0kb late RNA"
88511
/note="TATA: TATAAGA"
complement(88514)
/note="TATA: TATAAT BL-L3 early promoter before BLRF3,
gives 1.0kb early RNA"
88547..88855
/note="BLRF1 late reading frame"
/codon_start=1
/db_xref="PID:g1334865"
/translacion="MGKVLKRPFAKAVPLFLAATWLLTGVLPAGASSPTNAAAASLT
EAQDFSYTCNADTFSPSLSFASINALLVLVLIATAYLMAYCENKRVNLLTD
"
promoter
88863
/note="TATA: TATTTAA BL-R2 late promoter before BLRF2,
gives 0.6kb late RNA"
88925..89413
/note="BLRF2 late reading frame, 2 NXS/T"
/codon_start=1
/db_xref="PID:g1334866"
/db_xref="PID:g1334866"
/translacion="MSAPRKVRLSVKAVDMDSMDMAARLARLESENKALKOQYLKRG
ACASTSVSPAPVPPPEPLAROREVMITQATGSLASQAMKIEDKRVKSYDVTTRN
EMENILQWLIRIQVSMGAKGSPGCGTIPRSDNPNAIRARRSRSGREAKKVI
SD"
89412
/note="polya signal: AATAAA, end of 1.0kb and 0.6kb late
RNAs"
complement(89425)
/note="polya signal: AATAAA, end of 0.7kb early, 2.2kb
late and 2.8kb late RNA"
complement(89430..92153)
/note="BLRF1a, late reading frame, gp350 membrane antigen,
36 NKT /S (Hummel et al, 1984; Biggin et al, 1984; Beisel
et al, 1985)"
/codon_start=1
/db_xref="PID:g1334869"
/db_xref="SWISS-PROT:P03200"
/translacion="MEALLVCQYTIQSLIHLTGDEDPGFENVEIPEPFYPTCNCTA
DVNVTINFDVGKKHOLDLDFGLTPHTKAYOPGAFGSEENATNLFLELLGAGEL
ALIMRSKLPINTTGEQVLSLEVDVYFQDVFGTAMGCHHAMQNPVYLIPETVYI
KMNCNSNTITAAVRAOGLDVLPLSLPTSADQNSFVSKTMEGNEDIEIMDGI
SOVLPGDNKFNITSCGESHVPSGGIILSTSPVATPIPGCTAYASLRTPRVSRLG
NNSLIFYSGNGPKRAGSDICIQSNITVFDEILPASQDMPNITDITYGDNATFDT
MVTSEDANSPNVTATFAMWPNNTETDFCKWLTLSGSGCENISGAFASNRFDIT
VSGIGTAPKTLITRTATNATTTTHKVIKFSKAPSTSTPLNTGADPTTGLS
STHVPNTLITAPASTGPTVSTADVTSPGATVSPGSPMDNGESKADMTS
PSTVPTPNATSPTPAVTTPPNATSPGATVSPGATVSPGATVSPGATVSPGATV
PNTSPITLGTSPSTSAVTTPPNATSPITLGTSPSAVTTPPNATSPITLGTSPSA
ATNLTGISTPTVVTISOPKNAISAVTGGNITSSSTSMLSLSPSSPELSTSD
NSTSHMPLITSAHPGTGENITQVTPASISTHVGSTSPAPRGITSOASGNGSTST
KPEGVNVTGTPONATSPQASGOKTAVPTVSTGKANSTGKHTGHGARTSE
PTTDYGDSTTPRRYNATVILPSTSKLRPMWTFSPVTTAQAATVPVPSQPRF
SNLSMLVLQNASLAVILLLLVMADCAFRNLSTSHYTTTPYDAETVY"
complement(89430..92153)
/note="BLRF1b, late reading frame gp220 membrane antigen,
spliced form of BLRF1a (Hummel et al, 1984; Biggin et al,
1984; Beisel et al, 1985)"
/codon_start=1
/db_xref="PID:g1334868"
/db_xref="SWISS-PROT:P03200"
/translacion="MEALLVCQYTIQSLIHLTGDEDPGFENVEIPEPFYPTCNCTA

```

```

CDS
89430..92153
/note="BLRF1a, late reading frame, gp350 membrane antigen,
36 NKT /S (Hummel et al, 1984; Biggin et al, 1984; Beisel
et al, 1985)"
/codon_start=1
/db_xref="PID:g1334869"
/db_xref="SWISS-PROT:P03200"
/translacion="MEALLVCQYTIQSLIHLTGDEDPGFENVEIPEPFYPTCNCTA
DVNVTINFDVGKKHOLDLDFGLTPHTKAYOPGAFGSEENATNLFLELLGAGEL
ALIMRSKLPINTTGEQVLSLEVDVYFQDVFGTAMGCHHAMQNPVYLIPETVYI
KMNCNSNTITAAVRAOGLDVLPLSLPTSADQNSFVSKTMEGNEDIEIMDGI
SOVLPGDNKFNITSCGESHVPSGGIILSTSPVATPIPGCTAYASLRTPRVSRLG
NNSLIFYSGNGPKRAGSDICIQSNITVFDEILPASQDMPNITDITYGDNATFDT
MVTSEDANSPNVTATFAMWPNNTETDFCKWLTLSGSGCENISGAFASNRFDIT
VSGIGTAPKTLITRTATNATTTTHKVIKFSKAPSTSTPLNTGADPTTGLS
STHVPNTLITAPASTGPTVSTADVTSPGATVSPGSPMDNGESKADMTS
PSTVPTPNATSPTPAVTTPPNATSPGATVSPGATVSPGATVSPGATVSPGATV
PNTSPITLGTSPSTSAVTTPPNATSPITLGTSPSAVTTPPNATSPITLGTSPSA
ATNLTGISTPTVVTISOPKNAISAVTGGNITSSSTSMLSLSPSSPELSTSD
NSTSHMPLITSAHPGTGENITQVTPASISTHVGSTSPAPRGITSOASGNGSTST
KPEGVNVTGTPONATSPQASGOKTAVPTVSTGKANSTGKHTGHGARTSE
PTTDYGDSTTPRRYNATVILPSTSKLRPMWTFSPVTTAQAATVPVPSQPRF
SNLSMLVLQNASLAVILLLLVMADCAFRNLSTSHYTTTPYDAETVY"
complement(89430..92153)
/note="BLRF1b, late reading frame gp220 membrane antigen,
spliced form of BLRF1a (Hummel et al, 1984; Biggin et al,
1984; Beisel et al, 1985)"
/codon_start=1
/db_xref="PID:g1334868"
/db_xref="SWISS-PROT:P03200"
/translacion="MEALLVCQYTIQSLIHLTGDEDPGFENVEIPEPFYPTCNCTA
DVNVTINFDVGKKHOLDLDFGLTPHTKAYOPGAFGSEENATNLFLELLGAGEL
ALIMRSKLPINTTGEQVLSLEVDVYFQDVFGTAMGCHHAMQNPVYLIPETVYI
KMNCNSNTITAAVRAOGLDVLPLSLPTSADQNSFVSKTMEGNEDIEIMDGI
SOVLPGDNKFNITSCGESHVPSGGIILSTSPVATPIPGCTAYASLRTPRVSRLG
NNSLIFYSGNGPKRAGSDICIQSNITVFDEILPASQDMPNITDITYGDNATFDT
MVTSEDANSPNVTATFAMWPNNTETDFCKWLTLSGSGCENISGAFASNRFDIT
VSGIGTAPKTLITRTATNATTTTHKVIKFSKAPSTSTPLNTGADPTTGLS
STHVPNTLITAPASTGPTVSTADVTSPGATVSPGSPMDNGESKADMTS
PSTVPTPNATSPTPAVTTPPNATSPGATVSPGATVSPGATVSPGATVSPGATV
PNTSPITLGTSPSTSAVTTPPNATSPITLGTSPSAVTTPPNATSPITLGTSPSA
ATNLTGISTPTVVTISOPKNAISAVTGGNITSSSTSMLSLSPSSPELSTSD
NSTSHMPLITSAHPGTGENITQVTPASISTHVGSTSPAPRGITSOASGNGSTST
KPEGVNVTGTPONATSPQASGOKTAVPTVSTGKANSTGKHTGHGARTSE
PTTDYGDSTTPRRYNATVILPSTSKLRPMWTFSPVTTAQAATVPVPSQPRF
SNLSMLVLQNASLAVILLLLVMADCAFRNLSTSHYTTTPYDAETVY"
complement(89434)
/note="TATA: TATAAG"
complement(89567..90013)
/note="BLRF2 early reading frame (BLRF3 in Baer et al,
1984)"
/codon_start=1
/db_xref="PID:g1334867"
/db_xref="SWISS-PROT:P03199"
/translacion="MCPVQHPAQAPPAKQALEVPHQNGRLMSKRAPPKMOR
RPRPVAKRRRFRSPQOVRPTLPVVESTPOMEGQVOSPQITAVIQLRODRDTM
RPIIYLPALANGCPAGLRAHRLPQPKPCOSRORPSPDSQTSPC"
complement(90051)
/note="TATA: TATAACA BL-L2 early promoter before BLRF2,
gives 0.7kb early RNA"
complement(90062..90652)
/note="intervening sequence in gp220 gene"
90177..90639
/note="21 copies of 21bp approximate repeat"
complement(92192)
/note="TATA: TATTTAA BL-L1 late promoter before BLRF1a,b.
gives 2.8 and 2.2kb late RNAs"
92238..92581
/note="Exon in Bodescot et al (1986) RNA (spliced from
20763 to 92670)"
92243..92602
/note="BLRF3 reading frame"
/codon_start=1
/db_xref="PID:g1632791"
/db_xref="SWISS-PROT:P03202"
/translacion="MDKDRPGPALDNNMEEVSTSVYQVGSAGDMENVLTLSDS
SSEKEADAHLEPAOKGTKRKRVDHDAGSAPAPMLPPQPDLPGBEAILRRPDLRL
TILQIGAATVATSIPIA"
92670..95248
/note="Exon in (Bodescot et al, 1986) RNA from 92581, to
3' end"
92670..95162
/note="BERR1 frame, homology with BRRF2 and BERR1. A
fusion of BLRF3 with BERR1 encodes EBN-3a, latent cycle
gene. (Hennessy et al, 1986; Joab et al, 1987). Protein
sequence is in conflict with the conceptual translation"
/codon_start=1
/db_xref="PID:e25053"
/db_xref="PID:g1334871"
/db_xref="SWISS-PROT:P12977"
/translacion="RIDTRAIIDPFSSQISNTMTYIMATAIQAIIDRRRNASRD
QAWRLQTLAAGPMKGIQVASSMNYSTHQTPTVHQAATLGAGRGCHYTFAG
TFELPRTDPDRMLLVYOSVGNIVSCNRYISFPDYMAIHSLKIMEVLTLPOR
VSMERFAGTDFORDLSIKSFVDALGTSIOPWIDNPGSTTAQAMNAGFLRGA
GIDLLRTEGHVGAIGEMTEESDESDGDDDLPCIVSRGPKYKRPPIFIRLRLR
LILMRAGKRTQEGKEVLEKARSGTYTTPRPVYKRPVPOSETATSHSASVPRP
TILAAQAGATPLHEOHGNAPCVQAAPPTPLPVSPGDLREVSFSGKACAPVAP
AGGIVRWEESLQAAQAPAPVAPQHPMPEVPTVLALEVRVYKRPVAPAPK1AM
QCGEESGIRARERARERAPVAPVNPRESSQMSVBRRLALRAEAVOKQASVOPQO
LTVOPQOPEGLVPEOQMFPGAPSOVADVAVAGVPMOPOYDLPLIOPISQGA
PVAPILARSMGPVPVPAOTOPDIPDLTEPILINGASAHLRPOAMEGLIPQOMMF
GALISQSVRGAQOSQYDLPLIOPINHGAPAHFTHQPMPEBPVPEOMFGCAPS
QGDIVQOHQDALGALGYTLHGHNHGVSPVAVNYHLSQAAGFLPIDESGESSDTSE
PCBALDLISIGRPCQAPEMPVYQVEGQDADVEDLDSIHGRPRRTPEMPVQEGGON

```

misc\_feature  
92703  
/note="BAM: Bam HI L/E"  
complement(93161)  
/note="TATA: CATTAAT"  
93479  
/note="TATA: TATAGA"  
complement(93482)  
/note="TATA: TATAAAT"  
94208..94277  
repeat\_region  
/note="repeat type A"  
94281..94306  
/note="repeat type B"  
94307..94381  
/note="repeat type C"  
94386..94411  
/note="repeat type B"  
94412..94489  
/note="repeat type C"  
94490..94560  
/note="repeat type A"  
94571..94648  
/note="repeat type C"  
94649..94719  
/note="repeat type A"  
94896..94982  
/note="repeat type D"  
94983..95069  
/note="repeat type D"  
95221  
/note="polya signal: AATAAA"  
complement(95272)  
/note="polya signal: AATAAA"  
join(95353..95709,95788..98247)  
/codon\_start=1  
/product="EBNA3B (EBNA4A) latent protein"  
/db\_xref="PID:e276431"  
/db\_xref="PID:g1632792"  
/translation="MKRAWLSRAQADAGASGSDPPDYDGGDGNVTONGSEPSPEI  
GPELSASDDPQSGPVENLDAAREEPEEPEHGEHNGDDLDVHTPRPRVDPNP  
TQAPVIOLVHAYIDSMLOSDRLPLGSLFLEONLNIEFLWMCITVRRCQAKRP  
YNQRMKMLISSCRSMWNGRYTNLKNVSESGDNVPLVATLGCDDGTHTATY  
SAGIVQIPRISDONOKIETAFILARRASLSARVTLFELVSSGNTLYAIVIGLTK  
NRYSLIEFGWMLCKDHTIHEMROCTGRKAKPMLRAHPVAIPDDDLTMEIDL  
AYARGAMNIEARLPDDPIIVDDDESEIEEESDEEKSMESEKINPOLPPI  
TYGRPAVFDKSKDAKSTKCAIIVDFSVITKIEBHKKKAKARTQPRATESSQAP  
TVLQRPPTQOEPGPVPLSQARLEWOLPQPVYAVLHDESMQGVHSMQLDL  
LEKDDVEQEVNATILPVPQOPRAGRGCYFTGDLGIESDEPASTEVEHQLPLA  
PGDPLEIOPLTSPITSSSSASCAQTPWVOPSQTPDDPTKOSRPETAPRQW  
PMBLRPIPMKRLMOPIPENHPVGPPTHOPHOVEITPYKPTMAOIGHIPQPTGPA  
TMLRQMAPRTMOTPRAPRPMSPREPPYPRRRPRGAPRTPTPPPOVPPPRPRRGA  
PRTTPPOVPLPTMOLALRAPAQOGFTKOLHQLGKGKRPSTLKDALERDA  
AGQSPSGSSTDKIVQALIFYPVLOPIQVMGSGSPITMASAVTQATEXTERR  
GVGPMPTDIPPEKRAKIEAVTEPEMHGGSHPVILENVGQGOOQTLCCGTAKO  
ERMMLGIDIAVSSPSSSETSNDE"  
95353..95724  
/note="BERF2a reading frame"  
95725..98244  
/note="BERF2b frame, homology with BERF1 and BERF4. BERF2a  
and BERF2b are spliced together to make EBNA3B (EBNA4A)  
latent protein."  
complement(95819)  
/note="polya signal: AATAAA"  
complement(95853)  
/note="TATA: TATTAAT"  
complement(96276)  
/note="polya signal: AATAAA"  
97522..97698  
/note="3x60bp repeat"  
98323..98769  
/note="BERF3 reading frame"  
98364..98730

misc\_feature  
promoter  
misc\_feature  
repeat\_region  
mat\_peptide  
mRNA

CDS

/note="Exon in EBNA-1 RNA (Speck and Strominger, 1985)"  
join(98371..98730,98805..101423)  
/codon\_start=1  
/product="EBNA3C (EBNA 4B) latent protein"  
/db\_xref="PID:e276455"  
/db\_xref="PID:g1632793"  
/db\_xref="PID:g1632793"  
/translation="MESFEGCGDSROSPOENRDNVQTGEHDDPGPGPSSGASER  
LYPEESYRDQOPWGOSRGDENGMORIRRRRRRAALSGHLDIEDNPPLP  
ITPYTARNRDAACRAVKOSHLOALSNLIDSLDPLDILCFMARORLODRRPL  
VARGVGNHMLITSPSQPMGRTATRLTLTPVNRGADSLMLATGCONAKT  
LNTFSATVTPPAGPREQERYAREAEVRLRGKWRRYRIYDLIELGSLHINQ  
LQTEENLIDFVFEMGVMSSCNPNVANYFHTKIGNFXYPNNAIPNEPYARGI  
KEHIONAFERKAOIGLSMATGCEPGRDASTSGSEDEPGRGSSVELESSDELPY  
IDNMEPVQORPVMEFVRYPKARKRLPMTPTHTPKRTNVTSDRSKAEQSTPE  
RSGPSQSTVTPPARPTVEMVYILHQPVPYPAKPTTPPSRRRGACVYTD  
DVLEVIDETTEDSSVSQPNKRRKHODGFQSGRRQKRAAPVSPSOTGPVAGP  
PAAGPPAPAGPPAGPPAGPPAGPPAGRIIAPAGPPAGPIIVTPPSAPRIM  
APPVVMFMEROLPQSTGRKQCFEMEMAGREITQMOEPPSSHLSATQTPRPSM  
APSVCALSVMDAKAOPRESSHLSMSPPOPISEHOPREDDADLDSLHPDVAO  
PAROAYOGYQEPAPAROAYOGYQEPAPAROAYOGYQEPAPAGLOSSYGVGMPPT  
RSQHPCTYRHPMAWSQDPVHGTCQWDRAPHLPPQWDSAGHGDDQSOFPHLSE  
TGPRLQLSLVPLVSSAPSSWSPQPRAPRLPTPTPMPMLQDSMAVGCDSGTA  
CPSPFASDYSQCAFTPLDINATTPKPRVEESSHGPARCSQATAEAQETLSDNSEIS  
VFPKDAKQTDYDASTESELD"  
98731  
/note="DONOR: AAGGTGAGT donor"  
98805..101420  
/note="BERF4 frame, homology with BERF1 and BERF2b. BERF3  
and BERF4 are spliced together to make the EBNA3C (EBNA  
4B) latent protein."  
98805..99050  
/note="Exon in T4 CDNA (Bodesscot et al 1986). 99050 is not  
the end of the RNA."  
99126..102118  
/note="DEL: Deletion in Raj1"  
99443  
/note="TATA: CATAAA"  
100104  
/note="DONOR: ACCGTGAGT possible donor before repeat."  
100122..100304  
/note="10 x 15bp repeat"  
100528  
/note="DONOR: CTGTAGAG possible donor"  
100613  
/note="BAM: Bam HI E/e1"  
100665..100781  
/note="3x39bp repeat"  
complement(100860)  
/note="TATA: TATAACA"  
100919  
/note="BAM: Bam HI e1/e2"  
101426  
/note="BAM: Bam HI e2/e3"  
complement(101445..102116)  
/note="B2LF2 reading frame 3x NKT/S. 2.5kb late RNA  
traverses B2LF2, ends unknown."  
/codon\_start=1  
/db\_xref="PID:g1334876"  
/db\_xref="PID:g1334876"  
/translation="WVSFQVRYPLTATALVILLAFPLPRVGGGVAATAITW  
VPRPNVEVNDVDPDPVNFNKTAADEYGEVLELPHWTPLHFGVPPONTYRANCYIC  
NTRFETSYKGCCTFTTKKKTNNGCFQACALFCTTFTYGPDLPLVATRNLAIE  
SLWAGVYRVEGWNWISLDGTFRYVQIFSGHCTYVSKFSTVPSVSHSCFLKRCICVS  
QRNS"  
101690  
/note="TATA: CATAAA"  
101765  
/note="polya signal: AATAAA"  
complement(101786)  
/note="TATA: TATTAAG"  
101947  
/note="BAM: Bam HI e3/z"

misc\_feature  
mat\_peptide  
misc\_feature  
mRNA  
misc\_feature  
promoter  
misc\_feature  
repeat\_region  
misc\_feature  
promoter  
misc\_feature  
CDS  
misc\_feature  
promoter  
misc\_feature  
promoter  
misc\_feature  
mRNA

misc-feature complement(102098)  
/note="DONOR: CAGGTGAGG possible donor"  
mRNA complement(102126. .102341)  
/note="3' terminal exon of 0.9kb and 2.8kb early RNAs"  
promoter 102153  
/note="TATA: TATTAAT"  
complement(102156)  
/note="polyA signal: AATAAA 3' end of 0.9kb and 2.8kb RNAs encoding BZLF1 and BRLF1"  
promoter complement(102160)  
/note="TATA: TATTAAT"  
complement(join(<102210. .102338,102423. .102530, 102655. .103155))  
/note="BZLF1 reading frame, modified from Beer et al, 1984. Has two splices within frame. 2XNT/S. Immediate early gene which disrupts latency (Countyman and Miller, 1985), called EBI by Chevallier-Greco et al, 1986 and ZEBRA by Miller."  
CDS /codon\_start=1  
/db\_xref="PID:g1334877"  
/db\_xref="PID:g1632794"  
/db\_xref="SWISS-PROT:P03206"  
/translation="IMDPNSTEDYKFPDPYQVPEVQAFDQATRYODLGSPQAPL PCVIMPVLPPELPQGLTFAHYSTAFPGSMTSAPAPENAYQAAAPQLPVSDDITQ NOQTNOAGGEAPQEGDNSTVQTAAVVACFPAAGQGLADIGVPPAPVAPARTR KPOQPESELECDSELEIKRYKNRVAASRCRAKFKQLQHYREVAANKSSENDRLRL KQCPESLDVDSIIERTPDVLEHDLINF"  
promoter complement(102380)  
/note="TATA: CATAAAT"  
promoter 102415  
/note="TATA: TATAATAC"  
complement(102420)  
/note="TATA: TATAATAC"  
complement(102426. .102530)  
/note="Exon of 0.9kb and 2.8kb early RNAs"  
misc-feature complement(102504)  
/note="polyA signal: AATAAA, apparently not functional"  
repeat\_region 102581. .102652  
/note="semi-repetitive sequence, homologous to human c-fos 3' sequence"  
mRNA complement(102655. .103194)  
/note="first exon of 0.9kb early RNA encoding BZLF1"  
misc-feature complement(102918)  
/note="splice acceptor used in RZ fusion gene (Sargeant)"  
promoter complement(103231)  
/note="TATA: TTTAAA of BZLF1 immediate early promoter gives 0.9kb RNA"  
misc-feature complement(103256. .103311)  
/note="Upstream of BZLF1, homology to 106243 to 106188"  
CDS complement(103366. .105183)  
/note="BRLF1 reading frame, (immediate?) early gene, acts as transcription activator."  
/codon\_start=1  
/db\_xref="PID:g1334878"  
/db\_xref="SWISS-PROT:P03209"  
/translation="MRPKDGLIEDFLRLTPRIKQLGSLVSDYCNVINKETASVEI TLRSTYKICAKINEKAKHREMGGLMATINICNEFAIILRNRRVRRAENAGNDACSA CPTVMRYVLDLIVTDRFIOAPSNRVMIPATIGTAMYKLLKHSRYAVAYSVLIGY DRAAIMAGSKQVVELNMRKEGLSSKFAFCWVFVPLEEMFQIMVMSKRGHLL DDVADVRLALITLPRASISSHAGORSYGVLPACLLSTSKAAVEIPLVLVGADMD ELMGNDGASHTEARYSESGQFHAFTDELESLPTMTLKGASADCGDSSSSSDS GNSDTEQSEREAREARLAPRSKRTSRNPGOTPCPSNAAEQCPWIAVQESD ERPIFHPKSTPTPLPVYKRGKLRDSRGMLPKPEASALSIVEGREVCOPKIRP FHPGSPANRPDLASLAPPTPGVHEPVGSLTPAPVPOPIDPAPAVTPEASHLIEDP DEETSOAKVALREMAADTVIPQKEEAAICGCOMDLSPPRGHLDLITLLESMTEIDLN DSIPLPELNLITLDTFLNDECLLHAMHISTGLSIFDTSLF"

mRNA /note="exon in RZ fusion gene (Sargeant)"  
complement(104927. .104989)  
/note="BRLF2 poss. small 5' exon"  
promoter 105016  
/note="TATA: TATTAAT before BRRF1, possible promoter for 1.1 kb early RNA encoding BRRF1"  
promoter complement(105019)  
/note="TATA: TATTAAT before BRLF2"  
CDS 105182. .106114  
/note="BRRF1 early reading frame"  
/codon\_start=1  
/db\_xref="PID:g1334877"  
/db\_xref="SWISS-PROT:P03207"  
/translation="MASSNRGNARPLKSLFHELYLKHYPVGDVHLNTIGVDCDLP PSHLTAQGLFLARVLAQVQOHLLEDDITVPIKILAYFLELSYSYSPKDEQRDI AEVLDHDKTRDGLDRLWALLRKLDQDHHASAVNVLMPGSDYATNSLODYQISIG MKRYIADVCSSGTASMSMTATRLHSQULMASGSPSEPRKARFFNQVLTVAALCK FRCIYTVYIQGSTATISQULHLEIKLCSWIIISQDGMRLFOHSRPLITLWESVAANO EYVDATILPDCAEYIDILKTKHYLENCSSAMQYK"  
misc-feature complement(105185)  
/note="ACCEPT: splice acceptor in 2.8kb early RNA encoding BRLF1 and RZ fusion gene (Sargeant)"  
promoter 105213  
/note="TATA: CATTAA"  
promoter 106110  
/note="polyA signal: AATAAA, 3' end of early 1.1kb RNA encoding BRRF1"  
misc-feature complement(106125)  
/note="DONOR: CAGGTAGA possible donor"  
misc-feature complement(106188. .106243)  
/note="Homology to upstream region of BZLF1"  
promoter complement(106213)  
/note="TATA: CATAAA"  
promoter 106243  
/note="TATA: TATAAA before BRRF2, possible promoter for 1.8 kb RNA encoding BRRF2"  
CDS 106302. .107915  
/note="BRRF2 reading frame"  
/codon\_start=1  
/db\_xref="PID:g1334879"  
/db\_xref="SWISS-PROT:P03210"  
/translation="WSCQQRGSYVILVEHLAGALTKMSDFITGQDVTLSGNIAVKI RDAINQPPGGDIVALLISLFLANMALPTSGROSQRDLIPAAVALTATNLGCTVP GEMSHKDTPEGLALAYTGKLMNVDSGCGCELOCGKGLKAIKGYELPRLIIPTRK OCSFVNLNMLVHKLVLRGHVOLAYARVITPEHFHILPDSDDAVFAETLLAALEF LNMFEVILKDYITDQSMISLAKRGHMAAGNPLPAPPELLRDLRYEAFRSDNHVYF TTGPIANTQFPBELGRVYVYIDSSICAASHYQDVTITGVACGPRPRFSALPPAPSE PQGTCSQUTSRGNSRRNLGQPEGTSFPAVPVPCVLTSLTASAKONGMGSLHLAK PEETSPAVSPVCPILASPASRSKQHCVTGSSQAPSPSSVAPVAPASLGDLEEEBS RESPSLPSKKGDGEFEAMLEADANDLEDVQREFSGLRVIGDEDEDGSEDFESDLDL SDSHDEGDEGGAGGGRSLHSYLSLV"  
promoter complement(106385)  
/note="TATA: GATAAA"  
misc-feature complement(106973)  
/note="polyA signal: AATAAA"  
promoter complement(107124)  
/note="TATA: GATAAAA"  
misc-feature 107457  
/note="BAM: Bam H1 R/f"  
misc-feature 107565  
/note="BAM: Bam H1 f/k"  
misc-feature 107914  
/note="polyA signal: AATAAA, 3' end of 1.8kb RNA encoding BBRF2"  
misc-feature 107942  
/note="ACCEPT: splice acceptor for EBNA-1 RNA (from 98730)"  
CDS 107950. .109875  
/note="BKRFL encodes EBNA-1 protein, latent cycle gene."  
/codon\_start=1  
/db\_xref="PID:g1334880"  
/db\_xref="SWISS-PROT:P03211"  
/translation="MSDEGPGTGTGNGLGEKGDTSPEGSGSGSPQRRGDNHGRGRC

[illegible]

misc\_feature  
complement(111830)  
/note="polyA signal: AATAAA"  
112471  
/note="TATA: TATATAT"  
complement(112476)  
/note="TATA: TATATAA"  
112620  
/note="BAM: Bam HI K/B"  
113876  
/note="TATA: TATTTAT before BBRF1"  
complement(113885)  
/note="TATA: CATAAAT"  
114204  
/note="BBR1 late reading frame, homologous to RF 54 VZV"  
/codon\_start=1  
/db\_xref="PID: g1334884"  
/db\_xref="SMISS-PROT: P03213"  
/translation="MFKNVDESSAGALSGSALIVHTPAVSRLFFELLOGKIVYGGQ  
TITVALNRPVGSKRVFTLTKRAISHCTYDDVLTDMNKEEACIQKRPWSDSCARF  
RESTEVSSTMRKLTIVNDLTNTIRVLSHRSYVSEYERVIDATGMVAWVKPITVO  
ELHSHIKSLRRCVCRELGHERTIRISIGTELYEKREKIEISINSTFIPOPTETVIEL  
PRSDVYVYVGGRRRLHLVLEPPAIPAGVTVPVORLOINEMCYRTLEHAKIOOL  
LNTATLAKIIVGSGRDMYKODILALENSQKRPKKLLMLVLRKLSNKRKISCTVTV  
BEFTITDASNNIVDKNRLEFGDPGETAAGQKLVKKNYSIVYKCLTDINGQFOINGLE  
RELTKIRRMSESDIKVSLGPGGNNPAPASAPAAVAASVVDILTGSTASAIKELFNS  
PSAYSAGAVSGHNSILNSEFVSQIIPSRREMTKDLTELMSELEFTEFKLIPVVDNGO  
RLYVSSDITISILGPEPTYVALSEVLETVATGLVETVATGLIETDELVRSRLAIYIE  
DLGRYGCASATSGDGHGIRQASRGDTEPDHAKSRAPDPGAGS"  
115948 .116784  
/note="BBR2 late reading frame, homologous to RF 53 VZV"  
/codon\_start=1  
/db\_xref="PID: e25024"  
/db\_xref="PID: g1632797"  
/db\_xref="SMISS-PROT: P29882"  
/translation="MASGRHQPQGTSLTMQKYSLVATPRVLVEVNRHAIATVATNV  
PEFYVARNDLIRDLRAHVAKARMTISQFCGVVALSDSDVDVHNIIPHVYSERI  
LYKPNVARDLMICMLASMTENAKPSIGLREVIYGRITLLHSCKNLSDIFLENGART  
LLSTVYKHDEEGEATDPGWNEGILSEFKLHKELKRAPSEARDLMOSLFTLSGMCL  
LAPSDVYCADLNKEEDANSGETFNLFYQDSILTRHPCQVTLQTLRKCKGSDPTVKI  
IP"  
promoter  
complement(116683)  
/note="TATA: GATAAAA"  
misc\_feature  
complement(116696)  
/note="polyA signal: AATAAA"  
complement(116784 .11786)  
/note="BBRf3 early reading frame, spliced to BBRf2. BBRf3  
contains a consensus nucleotide binding site; Protein  
sequence is in conflict with the conceptual translation"  
/codon\_start=1  
/db\_xref="PID: e25025"  
/db\_xref="PID: g1334887"  
/db\_xref="SPTREMBL: O04343"  
/translation="AFTQGVKDSDEASRLDRDVGEGATVARRHRIYKARSGFCLLM  
AIFQDILVGGCRHSGPFLVWMEAFSWTLDOLAAREADKAPPSHDHLLTLVLDLTR  
RLAGERRNRFMALPRAMLOLRRAAGRLSGSHVCLIDKGAAPAPQATREGLISPT  
AYREIMAFILDVISALHPGVTIPMETIRETDELMTVLSEF"  
116785  
misc\_feature  
/note="polyA signal: AATAAA"  
complement(117386 .117513)  
/note="Inton spliced out in RNA linking BBRf2 and BBRf3"  
complement(117515 .119080)  
CDS

```

/note="BBLF2 early reading frame, spliced to BBLF3;
protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/db_xref="PID:e25026"
/db_xref="PID:g1334888"
/db_xref="SWISS-PROT:P30118"
/translation="PDVLKGPVLRNSQTMETPAESVARVSSVTFYNTQTGRWMA
IHWGIVPIKREDVETLIYVACQPLGSLPEPVNAPSTTEINFLNRELRSGG
LIAMLADAEKDLDFSRDRRLSARVEDEGLFQPLFPAGVYVCSQSGDGR
DOOPVVDGSGSEMEGQCPNARHSESQGLDVIIRPRGDVTFYVETEDDPSV
PRDILRPVTEVDLYSSDQATGRCGRDARRRYSKILPEAGFESVLNMSVAGG
LTAFLRSIYASCANHRGTPLEPLDPELCPGSDQFPYVGFPLPHYGRAPA
HWHRAPHSSEGLLDLNLGVSGPLADALIGDARSQGRSLDLOQIWPRIKEINP
RVHCTGEGEGGEDETIVYGAEAATAIADATWMLYLEARCHELSARCAPVGTDPGG
GOARDQOTWLRALHRYGSTDTRALGLTYATVRVLLHMAADGLTWAVADEILGFV
APSAHSEEPPLAQ"
118981
/promoter
/note="TATA: TATAAA BBR1 late promoter before BBRF3"
119067
/promoter
/note="TATA: TTTAAA BBR2 late promoter ?"
119098
/promoter
/note="TATA: TATTTA BBR3 late promoter before BBRF3"
119108
/note="DONOR: AAGGTGAAT possible donor"
119137..120354
/note="BBRF3 late reading frame"
/codon_start=1
/db_xref="PID:g1334889"
/db_xref="SWISS-PROT:P03215"
/translation="MKSSKNDFFVYRTWKTLYVFMVMSAVNPITAMPNLAGYC
FENALVDGALNLTNYNLAHLPILYLPPEPFVITLVFADCAFIYVAGFEVAL
IKARKVSGLTLDASAVSAGSEPTFLAILKMSIOVEIOVSYKHFVLAQFEVLH
FIASVYACACVTRFSPVWVKAODNSIPQDELMVVFYKLVVNTLIGLAEELI
VSLSYFLAIGNSFVMDVAVNLAFLIPIFYXITLFWLASFLSHNKGVCYG
FIASILILPIVRYEAVFYSAKILHTVAINVAITIPICCYAMIRICRIFKSGROTD
YVPSFETVELESEPRPSPKTPSGRNRRSSSTSSSRTRRQRPVSTALVSSV
LPMTDSEELPP"
120260
/misc_feature
/note="ACCEPT: ATCTCTCCAGGT possible acceptor"
120358
/note="polya signal: AATAAA"
complement(120747..120974)
/note="BBLF1 late reading frame, possibly homologous to RF
49 VZV"
/codon_start=1
/db_xref="PID:g1334890"
/db_xref="SWISS-PROT:P03216"
/translation="MGALWSLCRRRVNSIGVDGIIINLYNDYEEFNLETKLIAEE
GRACGTNGCLEDEDESENDELFLFNKRPN"
complement(120764)
/note="polya signal: AATAAA, 3' end of 0.6kb late, 1.6kb
early, 3.0kb early RNAs"
complement(120929..122341)
/note="BGLF3 early reading frame, homologous to RF 48 VZV
and alkaline exonuclease of HSV"
/codon_start=1
/db_xref="PID:g1334891"
/db_xref="SWISS-PROT:P03217"
/translation="MADVDELEDPMEEEMSYTFAFLRSPETEBAVYRNIDRPPQFAM
RFVYLCLCKQIOEFSGETGFCDVSSLAQVENDSKGSPLSKTSYWLQAEQQRVL
CSYVEMTRGQSENLMDILRNQIISSSKLSYIKNGPTKEVPAISINHGFGVVA
FGLRCEBDYKDIYCKLICGDASANROGFEMISPTDGIQVSLDLCYNVSGDEILFT
DRSCYIEIKREFYLFESKSEFPIYPSYALKRQCSTFIRINSIARPTVEYVPG
RLPSEGDYLLQDDAENLKDVRKRKLGPGHDLAYDSLANSRGESLTYMTQPSNAG
RIGIKRVPYNTFINRHNITFOVLLQYITGVGYVHSGGKGRGRCSPRVNITAF
KRSPLDPATCTGLDLDLDAVSEIIPVAVLPVPLVDSVIRKTLSTAAGSMKAYADN
TDTAPWPSGTFADDESTP"
complement(121331)
/note="TATA: TATTAA BBL1 late promoter before BBLF1"
121669
/promoter
/note="TATA: CATAAAT"
121697

```

```

/promoter
/note="TATA: TATAAG"
121772
/note="TATA: CATAAAG"
122313
/misc_feature
/note="BAM: Bam HI B/C"
complement(122328..123692)
/note="BGLF4 early reading frame, homologous to RF 47 VZV;
protein sequence is in conflict with the conceptual
translation"
/codon_start=1
/db_xref="PID:e25030"
/db_xref="PID:g1334892"
/db_xref="PID:g1334892"
/db_xref="SWISS-PROT:P13288"
/translation="SGMRSVSRSLRPTSCDRPSSHURNMDVNNAAELSPTNSSSG
ELSVPEPREPQEAQGLKVTYADYTFPOHKHLKVINIDMDETLYVKLPENNTCDHL
PTTCYILGSRSGYAVYAHADNATKVLDSYTELHELMQMDIOIGKATAGCDKA
LYDISACTSCHAELMPQPRCSLDQYGHMHGSIETPLVRCROGLKDAYFLNRKGLF
HSDISPSNIVDFDTMNGKRLVLTDTGTSLSLDRNKMADVRLKSSKROLYRLYQ
REPEFIADTYKPLCLSKCYILRGAGHIPPBSACGPVGAQATLRLDLSGLSLYG
IMHLADSTHKITPYPNPDMGFRSDPLYLQFAAPVVLVLEYSOMNINLMDGLTSCG
BSPCDVVAEHMSQFLQMCBSLKRKRESYFNCRRREHPLPBLVALADDFPGRG"
complement(123506)
/note="DONOR: AAGGTGACT possible donor"
complement(123941..124939)
/note="BGLF3 reading frame"
/codon_start=1
/db_xref="PID:g1334894"
/db_xref="SWISS-PROT:P03220"
/translation="WENAVKADMPDPPMLARRYGCCLELALACODTEDEKRLVETPL
KSFLLVSNILPQDNRPMHEARSRSRVADDDVFSLSALELPLNRLPEMOFGQGW
SRMPSQPEMGMLCFEYFDGDLIRALNWKDEVIQALQILHNSQWTSLYVEDP
LPWMAALYGPESHCEERHCYAAAKRGAPLITLATTYPTQAINAEALHTRCVYA
ILYDRDKMGEDIPAPDPVSRNKAOKCLIPQRPITRNCILIAKHONKNOYK
RVPYIDPCLITGAEKRYMDAVNRYEASGTIVLYPYTDLSIVADMTYEDE"
complement(124117)
/note="TATA: TATAAA"
complement(124219)
/note="polya signal: AATAAA"
124938..125915
/note="BGR1 reading frame, homologous to RF 45 VZV and
spliced HSV gene (Costa et al, 1985). Spliced to BDRF1.
Northern blots in BGR1 detect 2.7, 2.6, 2.1kb late and
1.9kb early RNAs. 2.6, 2.1kb RNAs very weak."
/codon_start=1
/db_xref="PID:g1334893"
/db_xref="SWISS-PROT:P03219"
/translation="MLVASQGRLEENRNLQODSTOGCIGAGTSPSIMYTGAKSRR
WAHPLVGTIHASNLCPMLRAVCRYGRPVFADESLPMPGASPALHTPOYOMCL
LPBELDQTLQRLPPNLDESEALTEFKTSVSASARAILDPNLEKREHVTSLASFLSG
CYKHKPARLEAKQOVVLSFEFLISIKSLEITDMPFOSAGLEEMTEKLIIFK
OKASVFLIPRRHGTWIVAITISLINSUNVOIGYVFAHOKHVASAVTEIITDITS
PDSKRVENKERSTTFRHSKISSTWMCATCENNVAPDVSVLNCNA"
join(124938..125873,129215..130351)
/note="BDRF1 reading frame"
/codon_start=1
/product="probable DNA packaging protein"
/db_xref="PID:e276433"
/db_xref="PID:g1632798"
/db_xref="SWISS-PROT:P03219"
/translation="MLVASQGRLEENRNLQODSTOGCIGAGTSPSIMYTGAKSRR
WAHPLVGTIHASNLCPMLRAVCRYGRPVFADESLPMPGASPALHTPOYOMCL
LPBELDQTLQRLPPNLDESEALTEFKTSVSASARAILDPNLEKREHVTSLASFLSG
CYKHKPARLEAKQOVVLSFEFLISIKSLEITDMPFOSAGLEEMTEKLIIFK
OKASVFLIPRRHGTWIVAITISLINSUNVOIGYVFAHOKHVASAVTEIITDITS
PDSKRVENKERSTTFRHSKISSTWMCATCENNVAPDVSVLNCNA"

```







HHEMASIDTGLSTSSMTTPARVAATITDMGCIHTODFFSVPPAEAGNOQVNDITAKV  
 GAQRNGTLLRDPTTYLAGMTINWAGPGLHGQOATCEIIVTPATAVAALFQKSNSPRG  
 RAACVVCSENNOEVAEGLIDHSPRDAAYEYKSTYNPAWSQAGSDIMNYSSTROT  
 ABGSLSPCAFEAKKEILLRNNGLYNMYENSOIRGCHPATNTETVOCVVLATGTVF  
 LADPCFLOEAFPALSSASRALIDEMFVKQTHAPIHYGHYIIIEEVAAPRILIKENK  
 VVF"  
 misc-feature  
 /note="(133332)  
 /note="DONOR: AAGGTGGTT possible donor"  
 complement(133352)  
 /note="TATA: TATTTAA before BDLF1"  
 complement(133386)  
 /note="TATA: TATATAA"  
 135178  
 /note="polya signal: AATAAA"  
 135394  
 /note="TATA: TATAAGT"  
 136624  
 /note="polya signal: AATAAA"  
 136868  
 /note="BAM: Bam H1 D/c"  
 complement(137710)  
 /note="TATA: TATTTAA EHL1 promoter before BclF1, gives  
 4.5kb late RNA"  
 137857  
 /note="TATA: CATTAAC"  
 137911  
 /note="BclF1 reading frame"  
 /codon\_start=1  
 /db\_xref="PID:g1334907"  
 /db\_xref="SWISS-PROT:P25215"  
 /translation="MLAHNVYTRIPPCPFSGREARLKHEFFSWTFMLSPNNATL  
 REIRTAANLHHPHLVDLYHASPTFLRSAGLYRVCNCTLLENISIOCKA  
 GRPGLEIILISNGGSPASFOSSPTGSLRCLIVAAALPEVSGHOELSLRSR  
 SOGGDIDVSGSPARALVALRREDGAPKDPGLGPHGPRGPGAPKSDSESRDA  
 PPGIDSSFOASRLVPGVPGFLVNTNRYNTKVCSEPLTKMKVCVPLINNV  
 AKYVYKETAFTYSLEFTDGVGANLAINNISITSLFLAMTSLCFLPEVETIYPA  
 VSMNLTDLHGLENOSLVRENKSGVFWTNPESVSCDGLVSWFKATYATISVH  
 GQTEQHLREITPIVTHREAKISRIKNLFTLLRNISOQLVLRKLEGLDCA  
 LRLDPSNCINRIASEGLDFSKRSIASHNKRHECALGHRSHANVTKLVNREKTELD  
 ILGRNANFLTRCKHOVNLROSPITFLLRHIRRRLGLGRASVREITLLAHLRKTA  
 PIHCRDAOV"  
 138019  
 /note="BAM: Bam H1 c/b"  
 139352  
 /note="BAM: Bam H1 b/T"  
 139642  
 /note="BclF1 reading frame. Northern blots detect 0.95  
 late and 3.8kb early RNA; Protein sequence is in conflict  
 with the conceptual translation"  
 /codon\_start=1  
 /db\_xref="PID:g25011"  
 /db\_xref="PID:g1334904"  
 /db\_xref="SWISS-PROT:P30119"  
 /translation="NERLPFSMPTCKRPPSTAVMLCKQPEARFTHGAVHLPSQI  
 VHTHSPTLASALGLGENVPIPALFRASGLNVRSLMTNMRAPITISLARIILAPN  
 PYLEQOLVGMQDNGIPIVLFARPIEVKSGESNIIKASSOLIMEDSKLQAPFS  
 ASEHPASWESVYKRVIRDEGANTRTIRIDILEIPYVLSIOLSPKSLIKAPPP  
 PREPOTFDPATPYARLFIYDIOGVPKLGNAAPAAQISAEKRAEKLEQAKNVENTIF  
 LOHLYKHYVLSINLENTVOPEEAFETSPSPQISANETKRAEKLEQAKNVENTIF  
 CUNICSTSTLDPVVDVNNPINSIALEKTFLEFPSPSGTINRVRPATPIVRLCOGA  
 ELGTMAQFLGKIKVKEKGTMTLVLYLRLI"  
 complement(140902)  
 /note="polya signal: AATAAA, 3' end of 2.5kb late RNA"  
 complement(140916  
 /note="BclF2 late reading frame, encodes gp85, homologous  
 to Rf 37 VZV and glycoprotein H of HSV (gp111 of VZV)"  
 /codon\_start=1  
 /db\_xref="PID:g1334905"  
 /db\_xref="PID:P03231"  
 /translation="MQLCVCFLVILMEVGASISPVKHLIDIEGHASHYTIPTMELM  
 AVPGISPPALMREANVTEDIASMLNRKYLITKTSTGLTALAEPPDITAVSEGSQV  
 DSKVHPGVISGLNSPACMLASPLEKOLFTYIGTMLPNTRPHSYVYQILRCHLSYVAL

SINGKFOYTGAMTSKFLMGTYKRVTERKDEHVLISLVEKTKDLPDLRGPSYSLTS  
 AOSGYSVIVTTFEYHANYFNPVNLKDEMSRANVTMTASAYARVLOKIVLEMKG  
 GRREDELTEITLMEFVSAFNEFGEKNCVYLMRLAKGFEFLKJDIKDIIGI  
 CGAVVKKMSOAGREELAMALMATYKMEIGLJTEKREYELRLATVAPYAGVYSGI  
 ICGATSVLSAVNRRDLFOLPLHTVYRLEFICSHVYVLEDELRLNTTQCPNATAYOLLS  
 TALCSALIEGYLRGLAIGTESGLSPCYLSRPLTQDKLSMAPQATFDDOAVSN  
 AVDFGLGRSLEREDRDVMIIPAYKCVRLKVLMIIPILNVTFLISSDREVSALY  
 EASTYLISSLETLSPVINMKCSOGAVAEPRQIPKIQNTFTRQKSCIFCGFALLSYDE  
 KLETTETVITISOEVONSILSSNYFDFNLHVHYLLTLTNGVMEIAGLYEBRAHVIL  
 ATTYEFLAFALGIFLVHKIVMEFL"  
 140970  
 /note="polya signal: AATAAA"  
 complement(141286)  
 /note="polya signal: AATAAA"  
 142589  
 /note="TATA: GATAAAA"  
 142740  
 /note="BAM: Bam H1 T/X"  
 complement(143038  
 /note="BclF1 early reading frame, thymidine kinase  
 (Littler et al. 1986). Weak homology to Rf 36 VZV and HSV  
 thymidine kinase. 4.0kb early RNA presumably encodes the  
 TK. Also a 2.2kb late RNA here."  
 /codon\_start=1  
 /db\_xref="PID:g1334907"  
 /db\_xref="SWISS-PROT:P03177"  
 /translation="MAGFPGKAGPGGCMRCKCOEDESPENERENFYAIEDFAPSVL  
 TPTGSDSGAEEDDGLGYOPTHMPPLNAPGLSGERVPCROQAVTNTGNSGSRH  
 TSCPTLPAGQAPAPAPANOKPATPPKRSRECGSKRPDPFSWRKTSCEGAGDST  
 SRSFMYOKGFEGLGLDLSDSDCESDESREKSPSSHALQKNGKGRPSGLEH  
 LAAHREKSKLSHAQOLKRLSGVANNLNDADDTQAAQORRESRVPVLTILTN  
 HVPYKRPACSLFLEGAPGVGKTTMNLKAVFGDULIVPEPMRYTHVADCAIKAM  
 KNVTPARRGREDTSAEVLACQMKFTTTPRVLASRRSLVTESGARVAPENALIKH  
 RHLASVPEPLMLRISOLISDIOVLAFTAPDGTIYMKRIANVENNRKLRKG  
 IRHSGEADGAGIKSYNDVAVHYAVQYCAVAPADYKVCACGLTITTYVCHOSTPI  
 IRRSVAEKILKXKISITSYKVEYQVPRBAVLEVCLATRTTILYQFLVLDSEPOD  
 LPGCYETIYMOALKNPATRSQFFDAGLSKVISDERGRND"  
 complement(143310)  
 /note="TATA: TATTAAGA ECL2 late promoter before BclF2,  
 gives 2.5kb late RNA"  
 143608  
 /note="polya signal: AATAAA"  
 144791  
 /note="ACCEPT: TCTTTCGTTTTCAGG poss. acceptor before  
 BclF1"  
 144860  
 /note="BclF1 late reading frame, homologous to Rf 35 VZV.  
 Basic (core) protein."  
 /codon\_start=1  
 /db\_xref="PID:g1334906"  
 /db\_xref="SWISS-PROT:P03232"  
 /translation="MDPTRGICALSTHOLAKFHSLLPARKKAQKRAHLRCYSKILSLK  
 SMEQLASPLSLPQGPPTDFRFLFVITIGRRIACDVVALQPYPCYIVERTAMST  
 ANPOSVTRKAQRLBETADLCDANFLRISCPVLSOSLEVALVFNQSLRLTQV  
 EFPALGQRTPLSTTGLNLPLSRWDGALRARLDRPRTAGCHRRTHVGRKPSOLTA  
 RYPRGARAGRGRGKGOVAGVOCPGAOK"  
 144862  
 /note="BAM: Bam H1 X/Y"  
 144945  
 /note="DONOR: CAGGTAAAC possible donor at 3' BclF1"  
 complement(145135)  
 /note="TATA: TATAACA before BclF1"  
 145302  
 /note="TATA: TATTTAA before BclF1, potential promoter for  
 1.9kb early RNA"  
 145416  
 /note="BclF1 early reading frame, homologous to Rf 34 VZV"  
 /codon\_start=1  
 /db\_xref="PID:g1334908"  
 /db\_xref="SWISS-PROT:P03233"  
 /translation="MALSGLVLDIPARLPRTGPELMAAPSLRNSLRVSPERALEAER  
 EAERARSSRWRCAGVLNRLRLVELDLSIDHILARAEIIRQDLDAAVAADFGLSEMO  
 VNSPSTGGRSAAPAPSPAPQFTRLTGNAGYAVASISPTDPLMWAGSLAOTLLGLNY

GNINQWVPSGPGWYRTMSANMORRVFPKOLRGNTLFTNSVSLKMTFVYVILEGTO  
DFFSDVRHLPLDQALILSVATLILDGSSHOORPPARSEELILEGPGSLEKTIIDL  
KAKSPGNGNMLITSGNKEARQSLAPLNKQAAVPPGFAFNKINLNVFAGSLPETTAL  
NVPGAGRDRDLVYRIANOIFGEDVPPFSHQMNILVGLAALAILVYTLICTANIA  
EATRLHLSILPLQAMORRRKPRMAGSAGVAPVOTLRPAALFEIYVRYRPA  
ADPOASISLPGVILALEIKIMDQAPSHVAINLTKGKPDILFEIINOKLIFHPA  
AMHAATQRLAFEDGVGVALGRSPMLARELLEKQFSASDDYDLVTLTGLYASP  
VAPS"

misc\_feature  
/note="(146926)  
/note="polya signal: AATAAA"  
147167  
/note="DONOR: AAGTAAAT possible donor"  
147170  
/note="polya signal: AATAAA, 3' end of 2.4kb late and  
1.9kb early RNAs"  
147721  
/note="TATA: TATTTAT before BVRf2, potential promoter for  
2.1kb early RNA"  
147927..149744  
/note="BVRf2 early reading frame, N-terminus homologous to  
RF 33 v2v"  
/codon\_start=1  
/db\_xref="PID:g1334909"  
/db\_xref="SWISS-PROT:P03234"  
/translation="MYQAPSVYVCGFVERPDAPPDACIHLDPVLVKSQPLPKKPLPL  
TVEHLDPADVGVFGLYQSRAGLFSNAISITSGFLDLDSIYHDCDIASQRLPLRE  
PVEVLAHMLPSSLASLADIPOTIADGKLSFEPHVSICALGRRGTAAVGTDLA  
VWLKHSDESPSIAQIENDANAKRESGCPEDHPPLPLRLIKALIDAGLRNRVETL  
RODRGVANIPASVYLKASDAPDLQPKDALQSPSTDPATMLSGNAEGARACGGS  
AAAGDILISVPRRTFTLQTLNDNKPRTPTPYAAPLPPFSHQIATAPSTGPAG  
AAPAGGITSFEGGITYAGPAGDPPGAFILMDATYHPHPHAPPPAYFGLGELFPPPV  
PYGSHLADYVAPARSNKRKRDEDEGGGLPFGSDATLYRDIAGLSVVEL  
OHTQLALRETTLSGHTGVGVCPOQGCPTHSQGFOPHQSEYEVRYVPHPPPTS  
HOAAQOPPPPTQAPAEACVAESTIPEAGAAGNSGPRDITNQOPTTEGHHNGKVLV  
QASASGVAQSKETTPKAKSVSHLKSIFCEELNKRVA"

misc\_feature  
148007  
/note="BAM: Bam H1 V/d"  
148620  
/note="TATA: TATTTAA ECRI late promoter before BdfR1,  
gives 1.2kb late RNA"  
148707..149744  
/note="BdfR1 reading frame: this is the C terminus of  
BVRf2"  
/codon\_start=1  
/db\_xref="PID:g1334910"  
/db\_xref="SPTREMBL:O6541"  
/translation="MISGNAGEBATAACGSSAAAGDOLISVPRRTFTLQTLNDNKP  
ROTPLYAAPLPPFSHQIATAPSTGPAGAAVAPAGGYTSPGGITYAGPAGDPPAF  
ADAAHTYHHPHPPPAYFGLPGLFPPPVPPYGSLSLRADYVAPARSNKRKRDEE  
DEGGGLPFGSDATLYRDIAGLSKSVNELQHTQLALRETTLSGHTGVGVCPOQGC  
YTHSGPYGQFOPHOSYEVRYVPHPPPTSHOAAQOPPPPTQAPAEACVAESTIPE  
AGAAGNSGPRDITNQOPTTEGHHNGKVLVQASASGVAQSKETTPKAKSVSHLKS  
IFCEELNKRVA"

misc\_feature  
149115  
/note="BAM: Bam H1 d/1"  
149727  
/note="polya signal: AATAAA, 3' end of 2.1kb early and  
1.2kb late RNAs"  
complement(149758)  
/note="polya signal: AATAAA, 3' end of 1.0kb late, 1.5kb  
late and 1.8kb late RNAs"  
complement(149779..150525)  
/note="BILF2 late reading frame 11kxxt/S"  
/codon\_start=1  
/db\_xref="PID:g1334911"  
/db\_xref="SWISS-PROT:P03218"  
/translation="MTHVLILCCVSCVCAFFSDLVKENTYAHAGANVNLCSVPS  
NNSVSIIELEGRTYPGDGLPLAVATSNNGTHTNGYNSLLEWNGSNVTSLSLI  
PNVTLAHAGYTCNVLKNGSVAGVHCNVSAGEEDDQHANNTLQORHMLVTPIT  
IAPTLVSHSTSTSHRDRPVASKRPTLHKPVTLGPRPDPWRKKTWVWMLLLITCA  
VVAAPVLIILITISGLWLAGMRRKRWIP"

CDS  
complement(150571)  
/note="TATA: TATTTAG before BILF2. Potential promoter for

repeat\_region  
1.0kb late RNA."  
151236..151618  
/note="repetitive sequence 3x"  
151767  
/note="polya signal: AATAAA"  
complement(151780)  
/note="TATA: CATAAAA"  
152012..152013  
/note="DEL: B95-8 deletion with respect to Raj1"  
complement(152161..153099)  
/note="BILF1 reading frame, membrane protein, 3kxXS /T"  
/codon\_start=1  
/db\_xref="PID:g1334912"  
/db\_xref="SWISS-PROT:P03208"  
/translation="MLSTMAPGSTVGTIVANNSTVATEDACTKSYSAFISGTSLL  
VLLIITLAGILFIIFVRLVHMDVWVALLIELLWVIGKMIQIESSTGCLITON  
NMFLGMSVWTHIGMALIEKTLAIPSPRKRSRHRVNCYLKMGVCTVLLITLIT  
MGPDANLNRGPNACREBPTKGMHTAVOGIKRAGCYLIAAVLVILVITIKLIRKIFG  
RKPRILCNVETFGLICAFSFMILSLPLFELGAGSIFGPDCTESLVARYYPGPAACIAL  
LLIILYAMSFSHRMDLSLKQVYVTAIFYRRVPSQSY"

promoter  
152230  
/note="TATA: CATAAAA"  
153259  
/note="polya signal: AATAAA"  
153637  
/note="HPN: 22bp 2-fold symmetric"  
complement(153690)  
/note="DONOR: AAGTGAGG possible donor"  
complement(153699..156746)  
/note="BALf5 DNA polymerase (early), homologous to many  
DNA polymerases, CMV HFLf2 and RF 28 v2v, 4.5kb early RNA  
apparently encodes BALf5, RNA ends unknown."  
/codon\_start=1  
/db\_xref="PID:g1334913"  
/db\_xref="SWISS-PROT:P03198"  
/translation="MSGGLFYNPFLRPNKGLKKPKDEKYLRLIPKCFQTPGAGVVD  
RGPOPICFYQDSLIVVGGDEDEGCKMMWRQAEQVAREADTHGSPDLFHYDILET  
VYTHKCAVIPSIDKQGYVPCGLIVILKRRADGASVCVNFEGQOAYVYASAPGLD  
VEFAVIALKASTPDRRTPCRVSYSEKVTNRSTMGNGHAGVDYHKLITLSPNSCHVAT  
WLDKHKGRCIIPANNADATRRVYLDNDFYTFGWTSCRAIPRIORHRSYALELDCVIG  
DLSVRDESSWSPYQALADFDICLGBEGFPITNEDGRLILOJISCVLMSGEEAGRIAR  
ILLTGTCEIDIEGVEYEFPSBLMDLYAFOLIRDSVITVIGYVANAIDWPYILDIRA  
RHYSINPASTLGRIRAGVCEVRRPHDACKGFLRANTKRVITGLPIIDMYAVCRDLIS  
LSDYKLDIVARHLGLAKKEDVYKKEIPEIFAAGPERKRGVACVODSALVMDLHNF  
VIHVEVAAELIAKTAHIDPCRRVLDGQOIRVCSLLAAQKENTLIPMSASDSDGYGA  
TYIOPLSGFYFNPVLYVDRASLYPSITQAHNLCISMTPTGEEHRLAGLRPGEDIESF  
RLUGGYTHVKNHVESFLASLUTSWLAKRAIKILKALCEDPDRORTLIDKQOLATIK  
TCNAVYGFVANGDLEPCLISIAETVILQRTMLERAKAVEALSPANQALAPSDAM  
APLNEGQGLRIVYIGDLSLFIICRGSESEETLRFADALAAHTRSLSFVAPISLEAKT  
FSCMLITKRRTVGYLTDKTLMKGVELRYTACKRVOTRCRRVLDVLAADRVEKA  
SILSHRPQESFTQGLPVGFLPVIDILNAYNDIRGRRVPMGEICSTSTSKLSAYK  
STOMPHLAIVYQKFEVERNEBLPOIHDRIOYVPEYPPKGVGAKRTEAEDPAYAERHGV  
PAVADHYFDKLQGAANILQCLEFNNSGAISVLONFTARPP"

misc\_feature  
154747  
/note="BAM: Bam H1 I/A"  
complement(156707)  
/note="polya signal: AATAAA, 3' end of 2.5kb late (9B) RNA  
and 1.8kb late RNA"  
complement(156749..159322)  
/note="BALf4 late reading frame 9kxxt/S homologous to HSV1  
glycoprotein B (Pellet et al, 1985), CMV HFLf1 and RF 31  
v2v (gpi1)"  
/codon\_start=1  
/db\_xref="PID:g1334914"  
/db\_xref="SWISS-PROT:P03188"  
/translation="MTRRRVLSVVVLLAALACRLGAPQEPQAPPAVTVQPATRQOT  
SEPPFVCSHSGDLFFRESDDIOCSFGTRENHTEBLAWFDNIIIPYFKVRSYTKI  
VNIILYVNWYADSVTNRHEKESVDSYSTDOMDITVOCYNAVKATKQDLTFTVYDRI  
GVNTIVNKLCPKGLANGVRYVASQETIEAPWILMTYRTVYTYCTITDPMAKNSNP  
EPPEYTTTQOTYEMKSPFDGKAKKEIFHEADSFHRYRNKTYIDYIDNRGINTPGGERAF  
LDKGYTTLSMKLENRTATVCPLOHWOFTSDTIAETFGKSIHFVYDEGTSSEVNTTVGI  
ELPDAFCEIEQVNRKTMHEKYAYODRYGGEALTYFTTSGLLIWLAMPILPRISLAT  
VNLNELTPTTSSPSPSPAPASARGSTPAVAVLRRRRRDGNAATTVPVPAAPKSL

GTLNPAITVOIQEAYDSLROINMLCDLAKWCLCEKRONMVLREITKNPTTWSS  
 IYGAVAKRLGDVIVSQCPVNOAVITLRKSRVYSESMCSRLEVSFINDRK  
 TYEGQLDNEIETIKMTKEVCATISOYFQSGEIVHYNDYHFKTIEEDJLOT  
 FLSINTSLIENIDFASLEIYSRDEORANVEDLEGIPREYNFOKNTAGLRKIDMAY  
 SNGRNOVDGIGLMDSLGSGOSITNLVSTGGLFSSISYGFSEFKNPRGGLIY  
 LVAGVILVISTRTROMSQOPYOMITPGIDELAQOHASGEFGINPISTELQAIM  
 LALHEONQOEORRAAQRAAGPSVASRALQAAADRPPGLRRRYRHYDETAALLGEAETE  
 F"  
 158204  
 /note="TATA: TATAAAA"  
 complement(159312..161678)  
 /note="BALF3 reading frame: Protein sequence is in  
 conflict with the conceptual translation"  
 /db\_xref="PID:e24966"  
 /db\_xref="PID:g1334915"  
 /db\_xref="SWISS-PROT:P25939"  
 /translation="GRRGVILGPLLRGGQRRPDPGHCIQRDVRVGGSGLSGGSV  
 GMARGSPGQLYQGRAGQHRVAGLGRRAAGRRRLNGQTOAPSGHSPRTGLETPG  
 AAMSGLLAAAYSOVYALAVELSVCTRLDPPRLDVAAYVRNAGLAEALILPRLRQ  
 NDRACSAISLELVHLENSREASALALPGKGTVPVLRTPSVAYSEVEYGGHKVY  
 SLCLINDIEMKRIINSVFQMSHTMGLESERLADLLGRGVSPIIDRLYITVSP  
 CMRVGELMWLPNHNPNSTAGTGHVSCNHLAVPNPVPVSGLPFENEVRQAGIILFA  
 EKRARGCGPEGAPGPGRPBEGATRALDTYVNFSTVPPVPAELSLILVNSGKAI  
 GATQGGGGHSHSLALFAERRLATVRGACEALAGARLTHLEDAVAPATERLFCG  
 GYVSSGDVAEALKADCAAFTHAPQYRAILOKRNEDYTRLNAMQRLGGEESAE  
 SPEVRPAGAREPGPGALDALKRKEQYQVQVTEGLAKQSCIAQOSETLTLCL  
 RMGDVYVVELARHNEILYBRAVSGPWEDEBRBEGAEFNSKYIKTHLETOIJSSE  
 HIAHTHSIVTFTITGLAESGLEPPPSVALACCAAGLTPQKAFSLIMPGE  
 PSDWIETSFNSFVPGGSLASSQOILCRALREVLIVLSLNTKMGSLILRRDAVS  
 PGQALPPDGLVLYDSORPLILLYKRGWFKDLYALLYLHLQMRDUSA"  
 complement(159370)  
 /note="TATA: TATTTAA ECL1 late promoter before BALF4,  
 gives 2.5kb late RNA"  
 160966  
 /note="polya signal: AATAAA"  
 160990  
 /note="polya signal: AATAAA"  
 complement(161013)  
 /note="polya signal: AATAAA, presumed end of 3.9kb early  
 RNA"  
 complement(161384..164770)  
 /note="BALF2 early reading frame, homologous to RF 29 VZV  
 and major DNA binding protein HSV. 3.9kb early RNA"  
 /codon\_start=1  
 /db\_xref="PID:g1334916"  
 /db\_xref="SWISS-PROT:P03227"  
 /translation="MGAQTSSEDMGSGSQDPGCGIYFFPLATPIREVAITLCTGYA  
 GHRCLIVPLCGITVEPGFSINVALRRRDPDNGGLRATSYHDIYVFHNAHVPI  
 FEGPGLALGETREVEFGYASALPRESSKPGDFPEGLDPSAYLGAVALITEAFER  
 LYSGLNALISLKOEVAGOSAVRPLDYKEVPEGVOLRQFNLDSRCHEALY  
 TGLAOLARVRVAGLVELLEKOSLODQAKVAKVPLEKPEASTISHPDGSLMVDVA  
 ACELAVSYAPAMLEASHETPASLIMYSMPLEADCEGBARAAHARNAISLAPVSTQ  
 IFTANSVLYISGSKSTGGKRESLFSNFTMHGTLIOBEGTMDCRPCSTGSGWGPV  
 TGINPGENVAVELHYAASPSNLNAYAYLLOQOQKSSLLPVPETGLSYVAGAAS  
 PMCSLCEGRAPAVCLNLFRLRDRFPVAYTORDPYVISGASGSYNEDFLGNFN  
 FIDKEDGQRPDEPRRTYVQNLNLERSLGIDAEGLKLEPHGPRPVFKKQY  
 DAADVAEVOGMNMAKNNTIYKDLVKSCYHVMQSCNPPAOPACPIETLOFYSLLT  
 ILDQISLPCMVCYNDNMGSGSPPEMLKGHYDCTNPFSLADKGVIAKAKVYH  
 GEPTCDLPDDALAQRYGRRLVPRSKVLMDLPNIKIKRNVFTGENAALDNSTI  
 KSTTRRENYIINGPYMKFLNTHKTLTPDKLSSLYIMHNSRRRSVSPVSGAABEY  
 SDLALFVGGSSRAHEESNVIDVYVGNLYTAKQNLNAILKACQTOFYISLLOGLVP  
 RTOVSAPARDYVGHTRAVESAAYAEATSSLATYVCAATDLSOVKARPVYTLV  
 VTINKYVGNNGNOIFQAGNLGYMGGRVBNLLOAQAGIRQAGSGSMRKKEVNT  
 PTGLITVYKRRQOATYTYEINIRAGLEAIIISQKQEBQCPVYCNLYDANGECASIT  
 RDBAEYILGRFSVLADSVLETLATIASGSIEMTEAARDPLGEGWGPGAQDNFISV  
 AEPVSTASQASAGLLLGGGGGSGGRRKRLATVLPGLLEV"  
 163978..166635  
 /note="DEL: deletion in Raj1"  
 complement(164814)  
 /note="TATA: CATTTAA before BALF2, presumed promoter for  
 3.9kb early RNA"  
 complement(164851)  
 misc\_feature

CDS  
 /note="polya signal: AATAAA"  
 complement(164855..165517)  
 /note="BALF1 early reading frame, 0.7kb early RNA"  
 /codon\_start=1  
 /db\_xref="PID:g1334918"  
 /db\_xref="SWISS-PROT:P03229"  
 /translation="NMIAIALDSPHFGLSAYTILPPRFYHISLKPVSMDPELTRP  
 TDSVFVTRPEAVANVAPSPDDKVAESSTYFMFAMVAVFTDEKDESLPALVLCRLIKA  
 SLRKDRILYAEIACRTADIGKDTVRILITSVRAVAYNDYDWSRLRVLYCTVVA  
 VRNYLDLHKSAAVLAIAHYLALYRLRMFARLGRGMRSLRQFPYTMALSLTDPLK  
 SL"  
 165442  
 /note="TATA: GATAAAA"  
 165466  
 /note="TATA: TATAAGA EDRI early promoter before BAR1,  
 gives 0.8kb early RNA"  
 165504..166169  
 /note="BAR1 reading frame"  
 /codon\_start=1  
 /db\_xref="PID:g1334917"  
 /db\_xref="SWISS-PROT:P03228"  
 /translation="MARFIAQLILLASCVAAQAVTAPIGERYTLN SYMRVSLGPEI  
 EYVAPRLGCEOVULIGRMHNDVIFTEMPPRGFDIHRSNTEFLIYTAANTISDQNY  
 LCMKRGTEVYTKQEHLSYVKPLILSVHSRSQFPDFSVLTIVCTYMAFPHVQWILM  
 PECEVAPAPTAANGVMEKDGSLSVAVDLSLPKPMHLPTVCGKNDKEAHGVYVSGY  
 LSG"  
 complement(165713)  
 /note="TATA: TATAAG before BALF1"  
 166165  
 /note="polya signal: AATAAA 3' end of 0.8kb early RNA.  
 Also 1kb late RNA in this region."  
 166469..166475  
 /note="TATA: TTAATTT"  
 166498..166916  
 /note="Exon 1 of terminal protein RNA"  
 166561..166563  
 /note="Likely initiator met of terminal protein"  
 166614  
 /note="BAM: Bam HI A/Nhet"  
 complement(166946)  
 /note="polya signal: AATAAA These 2 polyA sites are 3' end  
 of 0.8kb early, 2.5kb late and 2.5kb latent RNAs"  
 complement(166950)  
 /note="polya signal: AATAAA"  
 complement(166998..167303)  
 /note="BNLF2 reading frame: protein sequence is in  
 conflict with the conceptual translation"  
 /codon\_start=1  
 /db\_xref="PID:e25001"  
 /db\_xref="PID:g1334919"  
 /db\_xref="SPTREMBL:004362"  
 /translation="EATMRGRLAFYATIRSPFRMSRKNAKKEVPEDEPP  
 TMRPIQORLIRRALGCGVRPDADGICQRHPLPGLVSTKNFDLISLCELGKGC"  
 complement(167304..167486)  
 /note="BNLF2a reading frame"  
 /codon\_start=1  
 /db\_xref="PID:g1334920"  
 /db\_xref="SPTREMBL:004361"  
 /translation="WYHVERALLEQOOSACGLPGSSTETRPSPCPEDPDVSRRLRL  
 LVVLGVFLGILCLLI"  
 167320  
 /note="TATA: CATAAAA"  
 complement(167525)  
 /note="TATA: TATAAAA EDL2 early promoter before BNLF2a,b.  
 gives 0.8kb RNA"  
 complement(168163..166965)  
 /note="BNLF1 coding part of exon c of latent membrane  
 protein"  
 168399..168574  
 /note="5 x 33bp repeats"  
 complement(168966..169041)  
 /note="intervening sequence in BNLF1"  
 complement(168966)  
 repeat\_region  
 intron  
 misc\_feature

```

misc_feature
    /note="ACCEPT: TTTTTCACGCCAGT possible acceptor"
    complement(169041)
    /note="DONOR: TTGGTAGA donor"
    complement(169042..169128)
    /note="BNLF1 exon b of latent membrane protein mRNA"
    complement(169129..169206)
    /note="Intervening sequence in BNLF1"
    complement(169129)
    /note="ACCEPT: TCCCTTTTCCCCAGT acceptor"
    complement(169201)
    /note="TATA: TATTACA EDL1A late promoter, gives 2.5kb late RNA"
    complement(169206)
misc_feature
    /note="DONOR: TGAGTAGT donor"
    complement(169207..169474)
    /note="BNLF1 exon a of latent membrane protein mRNA"
    complement(169546)
    /note="TATA: TACATACG EDL1 promoter before BNLF1 gives 2.5kb latent RNA (LMP)"
    complement(170094..170631)
repeat_region
    /note="terminal repeat 1 538bp"
    170632..171154
    /note="terminal repeat 2 523bp"
    171155..171692
    /note="terminal repeat 3 538bp"
    171693..172231
    /note="terminal repeat 4 538bp"
repeat_region
    171693..172231
BASE COUNT 34054 a 52511 c 50755 g 34961 t
ORIGIN

```

```

Query Match          6.2%  Score 151.6; DB 17; Length 172281;
Best Local Similarity 53.2%  Pred. No. 1.7e-19;
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

OY 1500 CTGCTGCCCCGACAGCTGCTCCACAGGCTGCCCAAGGCTGCCCAAA 1559
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108853 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108794

OY 1560 CTGCGCGACGCTGCGACGCTGCGACGCGACGCGACGCGACGCGACG 1619
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108793 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108734

OY 1620 CACCTACAACCGCATTCCTACTACGCGCCCGACGCGCTGCCCAAGCT 1679
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108733 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108674

OY 1680 CTGCTGCGCTGCAATCCCAACCACTGTCACAGTACCACTACTGCTCC 1739
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108673 CTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108614

OY 1740 CCCCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108613 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108554

OY 1796 --GCCCGACGCTGCGCGACGCTGCGCGACGCGACGCGCTACTACTG 1853
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108553 CCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108494

OY 1854 AAATCCCACTACTGTCATACATGACGCTACTGCTGCGCGACGAGCA 1913
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108493 CTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108434

OY 1914 CCAATACCACTACTGTCATACATGACGCTGCGCGCTACTACTGCTG 1973
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108433 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108374

OY 1974 ACACCACTACTGTCATACATGTCGCGCGACGCTGCTGCTGCTGCTG 2033
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108373 CCCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108314

OY 2034 ATGACCCCAACCACTACTGTCATACATGACGCTGCTGCTGCTGCTG 2093
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

```

```

DB 108313 CTGCCCCCTGCTGCCCCCTGCTGCTGCTGCTGCCCCCTGCTGCTG 108254
OY 2094 AAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2145
    |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
DB 108253 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 108202

```

## RESULT 12

HS4B958RAJ/184113 bp DNA VRL 12-APR-1996

## LOCUS

HS4B958RAJ 184113 bp DNA VRL 12-APR-1996

## DEFINITION

Epstein-Barr virus, artificial joining of B95-8 complete genome and the sequences from Raji of the large deletion found in B95-8.

## ACCESSION

M80517 M75989

## KEYWORDS

9330330

## SOURCE

ORGANISM

Human herpesvirus 4 DNA.

Human herpesvirus 4

Viruses: dsDNA viruses, no RNA stage: Herpesviridae;

Gammaherpesvirinae; Lymphocryptovirus.

1 (sites)

REFERENCE

AUTHORS

Sample,J., Brooks,L., Sample,C., Young,L., Rowe,M., Gregory,C.,

Rickinson,A. and Kieff,E.

Restricted Epstein-Barr virus protein expression in Burkitt

Lymphoma is due to a different Epstein-Barr nuclear antigen 1

transcriptional initiation site

Proc. Natl. Acad. Sci. U.S.A. 88 (14), 6343-6347 (1991)

91296817

4 (bases 1 to 184113)

JENSON,H.B.

GenBank Curator Program

Unpublished (1992)

COMMENT

The B95-8 genome (V01555) has a large deletion in the right side of

the genome which has been sequenced in Raji (M3547). These

sequences have been joined to form an extended and more complete,

although artifactual, EBV sequence.

For features, refer to feature tables of V01555 and M3547.

FEATURES

SOURCE

1. 184113

/organism="Human herpesvirus 4"

/db\_xref="taxon:10376"

1. 152008

/note="B95-8 sequences (corresponds to 1-152,008 of V01555)"

152009..152012

/note="Overlap of B95-8 and Raji sequences at B95-8

deletion point (corresponds to 152,009-152,012 in V01555,

and 1-4 in M3547)"

153013..163839

/note="Raji sequences (corresponds to 5-11,831 of M3547)"

163840..163843

/note="Overlap of B95-8 and Raji sequences at B95-8

deletion point (corresponds to 152,009-152,012 of V01555,

and 11,832-11,835 of M3547)"

163844..184113

/note="B95-8 sequences (corresponds to 152,013-172,282 of V01555)"

BASE COUNT 36002 a 55824 c 54622 g 37665 t

ORIGIN

Query Match 6.2%; Score 151.6; DB 17; Length 184113;  
Best Local Similarity 53.2%; Pred. No. 1.7e-19;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

1500 CTCTGCGCCGAGAGGCTGCTCCCAAGCGCTGCCCAAGCGGTGGCCCAAGCGGCGTCCCA 1559  
108853 CTCTGCGCCGAGAGGCTGCTCCCAAGCGCTGCCCAAGCGGTGGCCCAAGCGGCGTCCCA 1559  
1560 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1619  
108793 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1619  
1620 CACTTACCAACCGCAATTCCTACTGCGGCGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCA 1679  
108733 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1679  
1680 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1739  
108673 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1739  
1740 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796  
108613 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1796  
1796 --GCCCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1853  
108553 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1853  
1854 AATGCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913  
108493 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1913  
1914 CCAACACACAGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC 1973  
108433 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1973  
1974 ACACGACAGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC 2033  
108373 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 2033  
2034 ATGACACCAACAGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC 2093  
108313 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 2093  
2094 AAGATGATGACATGACATGACATGACATGACATGACATGACATGACATGACATGACATGAC 2145  
108253 CTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2145

RESULT 13  
LOCUS HS4ULIR3 1150 bp DNA VRL 15-JUN-1990  
DEFINITION Epstein-Barr virus simple repeat array (1r3).  
ACCESSION J02079  
NID 9330445  
KEYWORDS repeat region.  
SOURCE ebv (epstein barr virus) from human.  
ORGANISM Human herpesvirus 4  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Gammaherpesvirinae; Lymphocryptovirus.

REFERENCE  
1 (bases 1 to 1150)  
Heller, M., van Santen, V.L. and Keilf, E.  
simple repeat sequence in Epstein-Barr virus dna is transcribed in  
latent and productive infections  
J. Virol. 44, 311-320 (1982)

JOURNAL  
MEDLINE  
FEATURES  
source location/Qualifiers  
1. 1150  
/organism="Human herpesvirus 4"  
/db\_xref="taxon:10376"

BASE COUNT 302 a 169 c 633 g 46 t

ORIGIN

Query Match 6.2%; Score 151.6; DB 17; Length 1150;  
Best Local Similarity 53.2%; Pred. No. 2.8e-19;  
Matches 347; Conservative 0; Mismatches 299; Indels 6; Gaps 1;

1500 CTGCTGCGCCGAGAGGCTGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTCCCA 1559  
908 CTGCTGCGCCGAGAGGCTGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTCCCA 849  
1560 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1619  
848 CTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 789  
1620 CACTTACCAACCGCAATTCCTACTGCGGCGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCA 1679  
788 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 729  
1680 CTGCTGCGCCGAGAGGCTGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTCCCA 1739  
728 CTGCTGCGCCGAGAGGCTGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTCCCA 669  
1740 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1796  
668 CTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 609  
1796 --GCCCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 1853  
608 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 549  
1854 AATGCCAATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1913  
548 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 489  
1914 CCAACACACAGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC 1973  
488 CTGCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 429  
1974 ACACGACAGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC 2033  
428 CCGCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 369  
2034 ATGACACCAACAGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTACGTAC 2093  
368 CTGCGCGACCGCTCCCAAGCGGTGGCCCAAGCGGCGTGGCCCAAGCGGCGTGGCCCAAG 309  
2094 AAGATGATGACATGACATGACATGACATGACATGACATGACATGACATGACATGACATGAC 2145  
308 CTCTGCTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 257

RESULT 14  
LOCUS D88733 3170 bp DNA VRL 19-MAY-1997  
DEFINITION Equine herpesvirus 1 DNA for membrane glycoprotein, complete cds.  
ACCESSION D88733  
NID 9214320  
KEYWORDS membrane glycoprotein.  
SOURCE Equine herpesvirus 1 (strain:HL1) DNA.  
ORGANISM Equine herpesvirus 1  
Viruses; dsDNA viruses, no RNA stage; Herpesviridae;  
Alphanherpesvirinae; Varicellovirus.

REFERENCE  
1 (bases 1 to 3170)  
Kirisawa, R., Kobayashi, T., Kawakami, Y. and Iwai, H.  
Submitted (31-OCT-1996) to the DDBJ/EMBL/GenBank databases. Rikio  
Kirisawa, Rakuno Gakuen University, Faculty of Veterinary Medicine,  
Department of Veterinary Microbiology, Bunkyo-dai-Midorimachi 582,  
Ebetsu, Hokkaido 069, Japan (E-mail: kirisawa@ca2.so-net.or.jp,  
Tel:+81-11-386-1112, Fax:+81-11-387-5890)

JOURNAL  
MEDLINE  
FEATURES  
source location/Qualifiers  
1. 3170  
/organism="Equine herpesvirus 1"  
/db\_xref="taxon:10376"

BASE COUNT 302 a 169 c 633 g 46 t

ORIGIN











CC sequences which direct proteolytic cleavage, adhesion, organisation of  
 CC the cellular matrix or extracellular matrix processing. Both naturally  
 CC occurring and non-naturally occurring MSF's may be characterised by  
 CC various combinations of alternatively spliced exons from this sequence,  
 CC with the exons spliced together in differing orders to form different  
 CC members of the MSF family.  
 SO Sequence 1404 AA;

Query Match 12.3%; Score 542; DB 1; Length 1404;  
 Best Local Similarity 30.4%; Pred. No. 6.6e-22;  
 Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

OY 93 PEPPAE-TTQAPATT---QAPTTQAPTTT---QAPTTQAPTTTQ-APTQ-AP 142  
 DB 331 PPKAEETTKGPAITPKKEPTTTPKEPSTTPKEPTTTPKSAPTTPEKAPATTTKSAP 390  
 OY 143 TTQ---APTQAPTTT---QAPTTTQ-APTQAPTTTQ---APTQ---APTQ 192  
 DB 381 TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 450  
 OY 193 PTTTQATTPATTPATTPATTPATTPATTPATTPATTPATTPATTPATTPATTP 252  
 DB 451 P-----APTTPKEPTP---TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTT 492  
 OY 253 DFDHLLPHDKYCNLFYQCSNGYTFEORCEGLYFNRYQRCDSAPNVECDGEISAP 312  
 DB 492 -----PTTPKEPAPTTTPK-----EPAPT 509  
 OY 313 VTGNEDEDDIDIGDLNCGCANFELDMLPHGNRCCKYQCVHGLVRRGAGTHSEF 372  
 DB 510 TTKE-----PS----- 516  
 OY 373 ELQOCDHIELVGLTPGSESEVEDEDACTGMYCPTPEIEMEPLP---NGCPADESIDH 429  
 DB 516 -----PTTP---KEPAPTTTKSAPT----- 533  
 OY 430 LHPHSDCGQYVLCVHGQIARPCGNLHFSPTATOSCESPVTAAGCOVECDSDNCTSTA 489  
 DB 533 -----TTKEPAPTTTKSAPTTPKEPSPPT-----TKEP 560  
 OY 490 APTA-----APTAPTAAPTAAPTAAPSTVVPATPATAPTAAPPTTAIPTAPTAAPTA 545  
 DB 561 APTTPKEPAPTTTPKAPAPTTTPKEPAPTTTPKEPA-PTTTKAPAP-TAPKEPATTTPKETA 617  
 OY 546 PTTAAPESPTT-----VTVPTAAPT-----APTAVPEIPTVT---SAPTAAPTAAPT 593  
 DB 618 PTTPKKLPPTPEKLAPTTPKEKAPATTPELAFTT---PEEPPTTPEEAPTTPKAAAPN 675  
 OY 554 AAPTAPPTAVPEIPTT-----VTSPTAAPT-----AAPAPNT----- 629  
 DB 676 TPKEBAPPTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPKELAP 735  
 OY 629 TTVTAPPTAAPTAPAP-----NTVTAPPTAAPT---AAPAPNTVTVPTAAPTAPAP 683  
 DB 736 TTTKEPTT-TSDKAPAPTTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKGTAPTTT 794  
 OY 684 AH-----APNTAAPTVTTS---APATPEDDIDP---PLNDP 720  
 DB 795 KKPAPKELAPTTTGTSTSDKAPAPTTTPKEKAPATTTPKAPAPTTTPKAP 842

RESULT 2  
 W48299  
 ID W48299 standard; Protein: 1721 AA.  
 AC W48299;  
 DT 17-AUG-1998 (first entry)  
 DE Cryptosporidium parvum GP900 antigen.  
 KW Open reading frame; ORF; antigen; GP900; cryptosporidium; infection;  
 KW antibody; prophylaxis; treatment; inhibition; retardation;  
 KW detection; diagnosis; human; 3' region.  
 OS Cryptosporidium parvum.  
 FH Key Location/Qualifiers

FT Domain 175..423  
 FT /note= "NINC domain"  
 FN W09806430-A1.  
 PD 19-FEB-1998.  
 PE 11-AUG-1997; U14104.  
 PR 14-AUG-1996; US-700651.  
 PA (REGC ) UNIV CALIFORNIA.  
 PI Gut J, Leech J, Nelson RC, Petersen C;  
 DR WPI; 98-159290/14.  
 DR N-PSDB; V20700, V20701.  
 PT Anti-cryptosporidium antibody - used to develop products for  
 PT detection, diagnosis, prophylaxis or treatment of Cryptosporidium  
 PT infections.  
 PS Claim 27; Pages 65-68; 89pp; English.  
 CC The sequence is that of the GP900 antigen which may be used  
 CC in the production of anti-Cryptosporidium antibodies. These can be  
 CC used for the prophylaxis, treatment, inhibition or retardation of  
 CC a Cryptosporidium infection in humans or in animals such as calves.  
 CC They can also be used for the detection and diagnosis of related  
 CC infections.  
 SO Sequence 1721 AA;

Query Match 10.3%; Score 453.5; DB 1; Length 1721;  
 Best Local Similarity 27.18; Pred. No. 4.4e-17;  
 Matches 175; Conservative 40; Mismatches 246; Indels 185; Gaps 20;

OY 99 TQOAPATTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 158  
 DB 270 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 329  
 OY 159 QAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 217  
 DB 330 TTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTTTNTTT 389  
 OY 218 PAATTPAATTPGVPATPAPWPAPWPAPWPAPWPAPWPAPWPAPWPAPWPAP 277  
 DB 390 TTTKPPTTTTATATTTTT-----ETESVIRKDEWCWL----- 423  
 OY 278 FEORCEGLYFNRYQRCDSAPNVECDGEISAPAPVTGNEDEDDIDIGDLNCGPANFE 337  
 DB 423 -----EKNGCEKAGATY-----VAVIGKDRGNIEMG 449  
 OY 338 IDWLLPHGNRCCKYQCVHGLVRRGAGTH--FSFELQOCDHIELVGLTPGSESEV 395  
 DB 450 AFTMIPND-----THVRRFRKVDGNTISVCRKRGAGKLEFP 488  
 OY 396 DVDEDACGTCPTPEIEMEPLPNCGRADFSIDLHPHSDCGQYVLCVHGQIARPCPG 455  
 DB 489 DRS-----LDFTIIPVAGHNS-C-SIIVGSGD-----G 515  
 OY 456 NLHFSPTATOSCESPVTAAGCOVECDSDNOC--TSTAPAPTAAPTAPTAAPTAPSTV 514  
 DB 516 KIHSPYSKSVSLSAIQSELEFNEVYCDTCKAYGAHISGYOTSDPYTTTAKRTT 575  
 OY 515 VPPATPAPTAAPVPPPTAIPPTAPTAAPTAPTAAPSPPTTVVPTAAPTAPTAAP 574  
 DB 576 T-----TTGAPGQPTT-----TTSPPSKPTTTT---TKATTTT 609  
 OY 575 EIPITVTSAPTAAPTAPTAAPTAPTAAPTAVPEIPTVTSPTAPAP---TTAPAPNTVT 631  
 DB 610 LNPITTT-----TOKRTTTTTPVGPKPPIATTTTTLKIVTTTITTKATTTT 659  
 OY 632 VPTAAPTAPTAAPNTVTAAPTAPTAAPTAPTAAPTAPTAAPTAPTAAPTAP 676  
 DB 660 TVPTTTTNTTKDEMTTTPPLPDIGDIEITPPIKMLDKTKRMAYDNGSLDSDNEP 719  
 OY 677 TAAPTVAHAPNTAAPT---TSAPATPEDDIDP---PLNDP 717  
 DB 720 IPGSGAGIADISNLFVQTHKSTLP-----IDMWGLPFP 757

RESULT 3

W31852  
 ID W31852 standard; Protein; 763 AA.  
 AC W31852;  
 DT 27-APR-1998 (first entry)  
 DE Mycobacterium tuberculosis 74 kDa protein.  
 KW Tuberculosis; mycobacteria; infection; diagnosis;  
 KM antimycobacterial; antidiabetic; vaccine.  
 OS Mycobacterium tuberculosis.  
 PD WO9741252-A2.  
 PN 06-NOV-1997.  
 PE 18-APR-1997; E01973.  
 PR 29-APR-1996; DE-017184.  
 PA (GBRB) GBR GES BIOTECH FORSCHUNG GMBH.  
 PI Esplita C, Honisch C, Moreno C, Singh M;  
 DR WPI; 97-549750/50.  
 DR N-PSDB; T93610.  
 PT New DNA and related proteins or RNA derived from M. tuberculosis -  
 PT used for diagnosis of mycobacterial infections, monitoring  
 PT vaccination and development of anti-mycobacterial agents  
 PS Claim 5; Fig 13; 55pp; English.  
 CC This novel 74 kDa protein is encoded by an open reading frame of  
 CC a Mycobacterium tuberculosis DNA fragment (see T93610) containing  
 CC polymorphic GC-rich sequences. Its amino acid sequence shows  
 CC a high proline content, but there is no homology to any known  
 CC proline-rich antigens of mycobacteria. Novel M. tuberculosis  
 CC proteins (see W31851-57) are claimed. These can be produced as  
 CC recombinant proteins, especially in bacterial, yeast, fungal or  
 CC higher eukaryote host cells, and used for diagnosing tuberculosis  
 CC and other mycobacterial infections in humans or animals. The  
 CC claimed proteins can also be used for epidemiological studies, for  
 CC monitoring vaccination, and for the development of vaccines and  
 CC anti-mycobacterial drugs.  
 SQ Sequence 763 AA;

Query Match 8.7%; Score 383; DB 1; Length 763;  
 Best Local Similarity 23.2%; Pred. No. 1.2e-13;  
 Matches 194; Conservative 40; Mismatches 299; Indels 302; Gaps 35;

14 VARREVSDAEKPNALHEPHDPXPAEQXLLPXEYDCTKFFYCEGLFIAPRDCAPGT 73  
 10 LAPLPAP 52  
 74 EKFSAQTCVHAALAGCTLPAPPAETQA-----PATTAQPTTAQPTTTQAPTT 124  
 53 -----PAPPKPKSKAPFPVPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 90  
 125 TTAQPTTTTAQPTTAQPTTAQPTTTTAQPTTTTAQPTTTTAQPTTTTAQPTTT 184  
 91 SPAPLPAP 141  
 185 QAPTTTAQPTTTTAQPTTTTAQPTTTTAQPTTTTAQPTTTTAQPTTTTAQPTTT 238  
 142 PVPNPP 201  
 239 WPPIGCLLP-NCPADFDIHLIDPKYCNLFYQCSNGTYFQRQREGLYFNPPYQRCDS 297  
 202 SPRRPAP 243  
 298 PANV-----ECDGELSPAPPTTEGNEDEDIDIGLLDNGCPANFEIDMLPHGNRC 348  
 244 PAVPIPPVPPPLPVNPKIPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 286  
 349 DKYVCVGHGNLVERCGAGTHSFELQCDHIELVGCITPGESESEVVDVEDACAGWC 408  
 266 -----PLNNHPPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 299  
 409 TEPIEMEPPLNG--CPADFSIDL--LPRESDCGGYIQCVCVHQTIA----- 451  
 300 VGVVPLAPLPNSHPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 348  
 451 ---RCFPGUL---HFSFPAQSCSEPVTTACQYFECDSNOCSTIAAPAPAP----- 496

DB 349 FCCRVCSGEVLAGALNPSRPS-RSPVLT-----TTPALPAPIDPLPLPLPL 394  
 QY 496 --TAAP-----TAAPTAAPTAAP--STVPPA-----TTPAATAP----- 527  
 DB 395 INTAVPIPLPVYALAPLPPLPLPLPLISGVPPAPPIPGKPKMTTTPPLAPAPPEP 454  
 QY 527 --VPPTTAIP-----TPAPTAAPTAAPTAAPESPTTVTPPT-----AAPTAA 568  
 DB 455 PVLPPGSCPESEKPNPPAPPEPEPKSSPALPAPAPAPAPAPAPAPAPAPAPAP 514  
 QY 569 PTTAVPEIPIYTSAPTA-----APTAAP-----AAPTAAATTAAP 605  
 DB 515 PRASMPALPPAPSPAPATRLCPPLPSPAPAPSPAPAPAPAPAPAPAPAPAPAP 574  
 QY 606 EIPITVTSPTTAAPTAAPAPNTVTPPTAAPTAAAPAPNTVTPPTAAPTAAAPAP 665  
 DB 575 NRPAP 626  
 QY 666 TTVTPPTAAPTAAAPPTVAHAPNTTAAPVTTTAPATTPEDDDIDPLPNDP 720  
 DB 626 -----PTPAAPKSRPALPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 670

RESULT 4  
 W43106  
 ID W43106 standard; Protein; 1664 AA.  
 AC W43106;  
 DT 16-OCT-1998 (first entry)  
 DE C. thermocellum O1PB protein.  
 KW Multimer; enzyme; complex; protein-protein interaction; dockerin domain;  
 KW cohesin domain; catalytic subunit; scaffold subunit; SdbA; synergistic;  
 KW cellulosome integrating protein; scaffoldin dockerin binding protein.  
 OS Clostridium thermocellum.

FT Key Location/Qualifiers  
 FT Domain 28..192  
 FT /note="cohesin type II domain"  
 FT Domain 207..363  
 FT /note="cohesin type II domain"  
 FT Domain 409..565  
 FT /note="cohesin type II domain"  
 FT Domain 607..763  
 FT /note="cohesin type II domain"  
 FT FR2748479-A1.  
 PN 14-NOV-1997.  
 PE 10-MAY-1996; 005854.  
 PR 10-MAY-1996; FR-005854.  
 PA (INSP) INST PASTEUR.  
 PI Beguin P, Leibovitz E;  
 DR WPI; 98-011569/02.  
 DR N-PSDB; T86623.  
 PT Cellulase proteins with cohesin or dockerin type II domains - useful  
 PT for potentiating the activity of multiprotein enzyme complexes.  
 PS Claim 7; Page 31-39; 60pp; French.  
 CC Multimeric protein, especially enzymatic, complexes are held together  
 CC by protein-protein interactions between domains designated dockerins  
 CC and cohesins, which are found on the catalytic and scaffold subunits  
 CC respectively. An example of such a complex is the cellulose degrading  
 CC protein complex from Clostridium thermocellum, known as the cellulosome.  
 CC This complex comprises around 15 proteins including endoglucanases,  
 CC cellobiohydrolases, hemicellulases, e.g. xylanases or lichenases, which  
 CC interact with a central "scaffold" protein designated the cellulosome  
 CC integrating protein (CIPA; see W43108). The catalytic subunits interact  
 CC with the CIP subunit via conserved 23 amino acid dockerin domains. CIP  
 CC has been shown to contain 9 copies of a cohesin domain.  
 CC The invention relates to the isolation of proteins binding to a novel  
 CC dockerin type domain found in the C-terminal portion of CIP. The new  
 CC domain is designated a type II dockerin domain (as compared to the type I  
 CC domain found on the catalytic subunits of the cellulosome). The type II  
 CC dockerin domain has some sequence similarity to the type I dockerins but  
 CC is unable to bind type I cohesin domains.  
 CC The sequence presented here is an example of a protein which binds  
 CC the novel type II dockerin domain and is the product of the O1PB gene.  
 CC The protein contains 4 type II cohesin domains in the N-terminal portion

CC of which the first domain (amino acid residues 28-192) is thought to bind  
CC C1pA. The novel type II dockerin and cohesin domains can be used in  
CC complexes, especially enzyme complexes, to potentiate their catalytic  
CC actions in a synergistic manner.  
SQ Sequence 1664 AA;

Query Match 8.5%; Score 372.5; DB 1; Length 1664;  
Best Local Similarity 23.5%; Pred. No. 9.1e-13;  
Matches 176; Conservative 66; Mismatches 325; Indels 183; Gaps 30;

QY 18 PEVSDAEKNPALHEPHDHPRAEQXLLPXEYDCTKFEYCEYLKFIAPRD-CAPGEFEK 76  
DB 825 PTPSD-EPTPS-DEPTPSDEPTPSDEPTPSER-EEPIPTDPSDEPTPSDE- 874  
QY 77 FSAQTCVHAALAGCTLPSPAETTQAPATTQAPTTTQAPTTTQAPTTTQAP 136  
DB 874 -----PTPSDEPTPSDEPTPSDEPTPSERPIPTDPSDEPTPSDEP 917  
QY 137 TTTQAPTTTQAPTTTQAPTTT-TOAPTTTQAPTTTQAPTTTQAPTTTQAP 192  
DB 918 TPSDEPTPSDEPTPSDEPTPSERPIPTDPSDEPTPSDEPTPSDEPTPSDE 977  
QY 193 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 251  
DB 978 PTPSD-PTPSDEPTPSDEPTPSERPIPTDPSDEPTPSDEPTPSDEPT-PSDEPTPSDEP 1033  
QY 252 ADEDIHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNRYQRCDSFANVECGEISAP 311  
DB 1034 TPSDEP-TPSD-----EPTPSERPEERIP-----TDTPSD-----EPTPSD 1068  
QY 312 PVTGEDEDEDIDIGDLNCCPANFIDMLPHGNRCDKYQCVGHNLVERRCGAGTHFS 371  
DB 1069 EPTPSDEPTPSDEPTPSDEPTPSER-EEPIPTDTPSD----- 1106  
QY 372 FELQCCDHIELVGCTLPGESEVEDDADCTGWCCTEPIEMEPPLNGCPADESIDHL 431  
DB 1106 -----EPTPSDEPTPSDEPTPSDEPTPSERPE-EPITDTPSD-----E 1149  
QY 432 PHESDCGOYLQCVHGOITARPCGNLHFSPATQSCSPVTAAGCOVECDSDNOCSTAP 491  
DB 1150 PTPSD-----EPTPSD--EPTPSDEPTP-----SDEPTPSERPE 1181  
QY 492 TAAPTAAPTAAAPTAAAPTASTIVPATPPATA-APVP--PTTAITPATTAAPTAAPT 547  
DB 1182 EPIPTDTPSDDEPTPSDEPTPSDEPTPSDEPTPSERPIPTDTPSDDEPTPSDE 1241  
QY 548 TAAPESPTTVTP-PTAAPTAAAPTAVPEIPIVTSAPTAAAPTAAAPTAAAPTAVPE 606  
DB 1242 PTPSDPTPSDEPTPSDEPTPSER-PEEP-IP-TDTPSDPTPSDEPTPSDEPTPSDEP 1297  
QY 607 IPTTVTSP-----PTAAPTAAAPTAVTVPPTAA-----PTTA-- 642  
DB 1298 TPEDEPTPSDEPTPSDEPTPSERPEEPIPTDTPSDDEPTPSDEPTPSDEPTPSDE 1357  
QY 642 -AAPNTTAAAPTAAAPTAAAPTNTTVVPTAAPTAAAP-----PTVAHA 666  
DB 1358 PTPSDPTPSDEPTPSERPEEPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 1417  
QY 687 PNTTAAPTTTTSDCATTPDEDDIDPLPND 716  
DB 1418 PTP-----TPTSKPTSTPAPTETEPTPSD 1442

RESULT 5  
W31855  
ID W31855 standard; Protein: 572 AA.  
AC W31855;  
DE 27-APR-1998 (first entry)  
KW Mycobacterium tuberculosis 55 kDa protein.  
KW tuberculosis; mycobacteria; infection; diagnosis;  
KW anti-mycobacterial; antibiotic; vaccine.  
OS Mycobacterium tuberculosis.

PN WO9741252-A2.  
PD 06-NOV-1997.  
PF 18-APR-1997; E01973.  
PR 29-APR-1996; DE-017184.  
PA (GBFR) GBF GRS BIOTRECH FORSCHUNG GMBH.  
PI Esplita C, Honisch C, Moreno C, Singh M;  
DR WPI: 97-548750/30.  
DR N-PSDB: T93610.  
PT New DNA and related proteins or RNA derived from M. tuberculosis -  
PT used for diagnosis of mycobacterial infections, monitoring  
PT vaccination and development of anti-mycobacterial agents  
PS Claim 11; Fig 16; 55pp, English.  
CC This novel 55 kDa protein is encoded by an open reading frame of  
CC a Mycobacterium tuberculosis DNA fragment (see T93610) containing  
CC polymorphic GC-rich sequences. Its amino acid sequence shows  
CC a high proline content, but there is no homology to any known  
CC proline-rich antigens of mycobacteria. Novel M. tuberculosis  
CC proteins (see W31851-57) are claimed. These can be produced as  
CC recombinant proteins, especially in bacterial, yeast, fungal or  
CC higher eukaryote host cells, and used for diagnosing tuberculosis  
CC and other mycobacterial infections in humans or animals. The  
CC claimed proteins can also be used for epidemiological studies, for  
CC anti-mycobacterial drugs.  
SQ Sequence 572 AA;

Query Match 7.6%; Score 335.5; DB 1; Length 572;  
Best Local Similarity 22.7%; Pred. No. 3e-11;  
Matches 170; Conservative 41; Mismatches 282; Indels 257; Gaps 32;

QY 20 VSDAEKNPALHEPHDHPRAEQXLLPXEYDCTKFEYCEYLKFIAPRDCAQTEFEKSA 79  
DB 1 VPAPKSKPASPAPRAPAP-----MPATPMEFP 29  
QY 80 QTCVHA-ALAGCTLPSPAETTQAPATTQAPTTTQAPTTTQAPTTTQAPTTTQAP 138  
DB 30 LPPVPDPISKETPPAPAPAPPT-PPAPVPDPVPLPVPVKNKIPAPAPPAVA- 85  
QY 139 TQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 194  
DB 86 LVAPRCPLPPLPNNHPAPAPAPVGVPLAPLNSIPAPAPAPVGVPLAPLPIGRGV 145  
QY 195 TTTQAAATPAT-----TPAATTPAATTPAATTPAATTPGVP----- 232  
DB 146 SVWKSFTLLSTFCRCVSGEVLGALANSPRSRSLTTPALPAPLPLPLPLPIN 205  
QY 232 -APTSAPVWPRIKCELLPNCAPDFDHLIPHDKYCNLFYQCSNGTTFQRCEGLYFNP 290  
DB 206 TAVPPIPPAPVYALAPLPP-----LAP-----LPISGVPAAPPIPG-- 247  
QY 291 YVORCSPANVECDGEISAPPVTEGNEDEDIDIGDLNCCPANFIDMLPHGNRCDK 350  
DB 248 WT-----TP-----PLAAPP-----EKTYPV-----LPGGSC-- 273  
QY 351 YVOCVHGNLVERRCGAGTHFSFELQCDHIELVGCSTLPGESEVEDDADCTGWCYPT 410  
DB 273 -----PSEKPNP-----APPE 285  
QY 411 PIEMELPNCAPDFSIDHLPHESDCGOYLQCVHGOITARPCGNLHFSPATQSCSPY 470  
DB 286 P-----PEPKSSPA-----LP-----PAPAPASMPASAVVPSP- 315  
QY 471 TAGCVFECDSNOCSTASTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 526  
DB 315 -----PIPAPAPAPASMPALP-PAPPSPATRILCPPLPSPPAANSPP 358  
QY 527 VPPTAIPPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 583  
DB 359 AAPAPPTPRLISANPCPVPVPAAPAPAPAP-APPELPAPAPDPTTPVANSPPAP 416  
QY 583 -APTAAPTAAAPTAAAPTAAAPTAVPEI-PTTVTSPPTAAPTAAAPTAAAPTAAAPT 640



DR N-PSDB; N60473.  
 PT DNA coding for Plasmodium falciparum antigens - expressing  
 PT poly:peptide(s) having antigenicity of RESA or FIRA antigens of P  
 PT falciparum  
 PS Disclosure: Fig 7: 55pp: English.  
 CC The inventors claim a novel DNA molecule which comprises a  
 CC nucleotide sequence corresp. to all or a portion of the base  
 CC sequence coding RESA (N60473) or FIRA (N60473). RESA and FIRA have  
 CC antigenicity suitable for providing protective immunity against  
 CC Plasmodium falciparum malarial infections.  
 SO Sequence 844 AA;

Query Match 7.4%; Score 327; DB 1; Length 844;  
 Best Local Similarity 22.5%; Pred. No. 1.3e-10;  
 Matches 183; Conservative 30; Mismatches 242; Indels 360; Gaps 34;

QY 96 PAET--TQAPATTQAPTTTQAPTTTQAPTT--TQAPTT--TQAPTTTQAPTT 143  
 DB 164 POETVEPTQEPVTEPTQEP--VTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 222  
 QY 144 TQAPTTTQAPTT--TQAPTT--TQAPTT--TQAPTTTQAPTTTQAPTT 188  
 DB 223 TVQEPVTVQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 282  
 QY 189 TQAPTTTQAPTTTQAPTT--TQAPTT--TQAPTTTQAPTTTQAPTTTQAPTT 244  
 DB 283 TQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 337  
 QY 245 LHPGCRAPDFIHLIPHDKNLFCYQCSNGYFQRCPEGLYFNPYQRCDSRANVED 304  
 DB 337 -----TVEHIDE----- 345  
 QY 305 GEISAPAPVTEGNEDEDIDIDGL--DNGCPANFEIDWL-LPHGNCRDKYQCVGHNL 359  
 DB 345 -----KKGSEEDNLSLSSETEFEKSKTKKKSSWLKFGGNKNDK-----KSK 389  
 QY 360 VERRCGAGTHFSFELQCDHIELVGCILPGGESEVEDVEDACTGWCPTPEIEMELPN 419.  
 DB 390 NEKK-----PSLESVKONADEQ-----K 407  
 QY 420 GCPADFISIDHLLPHESDCGOYLQCHGO--TIARPCGNLHFSRATQSCSPYTAGCOV 476  
 DB 408 EOPDTSQIS-----VNAODSVTIOEP-----TATQ--EPPPT-- 438  
 QY 477 FECDSDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 533  
 DB 438 -----GELTATQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 489  
 QY 534 PTPAAPTAAPTAAPTAA----- 551  
 DB 490 ATOEPSTQEHADKKASEGDNISLSRSETEEEKSKTKKKSSWLKFGGNKNDKSKN 549  
 QY 551 -----PESPT-----TVYVPTPAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 584  
 DB 550 EKKSLESVKONADEQKOPDTSQISVNAODSVTIOEPVTEPTQEPVTEPTQEPVTEPTQEPV 605  
 QY 585 TAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 630  
 DB 606 TTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 665  
 QY 630 -----TVTPPTA----- 637  
 DB 666 SEGDNISLSRSETEEEKSKTKKKSSWLKFGGNKNDKSKNEKKPSLESVKONADEQK 725  
 QY 637 -APTAAAPN-----TVYVPTPAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 684  
 DB 726 EOPDTSQISVNAODSVTIOEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 785  
 QY 685 HAPNTAAPTV-----TTSAPTPEDD 708  
 DB 786 QEPITQEPVTAQEPVTVQELTATQEPVTEPTQEPVTEPTQEPVTEPTQEPVTEPTQEPV 820

RESULT 9  
 P50073  
 ID P50073 standard; Protein; 907 AA.  
 AC P50073;  
 DT 17-OCT-1991 (first entry)  
 DE Epstein-Barr virus (EBV) outer surface protein.  
 KW Epstein-Barr virus; antigen; vaccine.  
 OS Epstein-Barr virus.  
 PN EP-151079-A.  
 PD 07-AUG-1985.  
 PF 26-JAN-1985; 400141.  
 PR 30-JAN-1984; US-575352.  
 PR 23-JUL-1984; US-633558.  
 PA (UYCH-) UNIV OF CHICAGO.  
 PI Kieff E, Tanner J, Hummel M, Belsel C;  
 DR WPI: 85-191978/32.  
 DR N-PSDB; N50114.  
 PT New fragment of Epstein-Barr Virus DNA - useful in vector to  
 PT express polypeptide for use in prepn. of vaccine against the  
 PT virus and for use in diagnosis.  
 PS Claim 2: Page 23-25; 26pp: English.  
 CC The sequence encodes an outer surface viral protein of EBV, used  
 CC to generate antibodies reacting with the surface proteins of  
 CC EBV-infected cells, and in the preparation of a vaccine against EBV.  
 SO Sequence 907 AA;

Query Match 7.4%; Score 325; DB 1; Length 907;  
 Best Local Similarity 23.2%; Pred. No. 1.7e-10;  
 Matches 161; Conservative 59; Mismatches 192; Indels 282; Gaps 29;

QY 100 TQAPATTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 156  
 DB 410 TNAATTTTTHKVFASKAPSTTSPLNTGTGADPTTTGLGSTHVPNTLNAPAST-GPTV 468  
 QY 157 TQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 201  
 DB 469 STADVTSTPTAGTSGASPVTPSPSPMDNGTESKAPMTSTSTPTPTPATSPPTAVT 528  
 QY 202 TPAATTPAAT--TPAATTPAATTPAATTP--GVPAPTSAPVWPPICELPLPGCCADDIH 257  
 DB 529 TP--TPAATSPVAVTP--TPNATSPILGKTSPTSATVP--TPN----- 569  
 QY 258 LLIHPDKXCNLFYQCSNGYFQRCPEGLYFNPYQRCDSRANVEDCEISAPVTEGN 317  
 DB 569 -----ATSP-----ATSP-----GKISPTSAVT-- 585  
 QY 318 EDEDIDIGDLIDNGCPANFEIDWLPHGNCRDKYQCVGHNLVERRCGAGTHSFELQOC 377  
 DB 585 -----TPPN----- 590  
 QY 378 DHIELVGCILPGGESEVEDVEDACTGWCPTPEIEMELPLNGCPADFISIDHLLPHESDC 437  
 DB 590 -----ATSP-----TLKTSPTSAVT--TPPNAT-- 613  
 QY 438 GOYLQCHGOTIARPCGNLHFSRATQSCSPYTAGQVECDSDNOCSTAAPTAAPTAA 497  
 DB 613 -----GPTVGETSP-----QANTNHTLGGTSTPTPVTS 641  
 QY 498 APTAAPTAA-----PTAAPTSTVPAPAPATAPVPTTAAPT-- 536  
 DB 642 QPKNATSAVTGQHNITSSSTSSMSLPPSSNPEFLSPSTSDNSHMLLTSAPHTGEN 701  
 QY 536 ---PAPNAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 586  
 DB 702 ITQVTPASISTHVVSTSP--PRDGTSGAGSPGNSSTSTKPGEVNVTGKPPONATSPQ 760  
 QY 587 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 632  
 DB 761 APSGOKRAVPTVITGSKANSTTGKHTTGAGATSTPTTDYGGDSTTPPRKRNATTYL 820  
 QY 633 PTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 692

```
Db 821 PPS---TSSKLRPRMTFTSPV---TTA---QATVPPTSQPRFSNLSMLVLQNASLA 870
OY 693 PVT-----TTSAPATPEDDDID 710
Db 871 VTLTLLLVADCAFRRLNSTSHYTPPYDDAE 904

RESULT 10
R80144
ID R80144 standard; Protein; 907 AA.
AC R80144;
DE 18-JAN-1996 (first entry)
DE EBY gp350/220.
KM EBY gp350; gp220; gp350/gp220; non-splicing variant; vaccine.
OS Epstein-Barr virus.
FH Key Location/Qualifiers
FT peptide 1..18
FT /label= Sig-peptide
FT 861..881
FT /note= "transmembrane region"
PN MO528488-A1.
PD 26-OCT-1995.
PE 13-APR-1995; U04611.
PR 18-APR-1994; US-229291.
PA (AVIR-) AVIRON.
PI Jackson WT, Spaete R;
DR WPI: 95-373802/48.
DR N-PSDB: T04821.
DR New DNA encoding a homogeneous gp350 protein - can be used for
PT preventing and treating Epstein-Barr virus-related diseases or
PT conditions
PS Disclosure; Fig.1; 61pp; English.
CC The donor and acceptor splice sites of the EBV gene encoding gp350/
CC 220 are mutated by replacement of native nucleotides by non-native
CC nucleotides, without altering the encoded amino acid sequence,
CC resulting in elimination of gp220 prodn. Recombinant homogeneous
CC gp350, useful in vaccines, is expressed in mammalian or insect cell
CC hosts.
SQ Sequence 907 AA;

Query Match 7.4%; Score 325; DB 1; length 907;
Best Local Similarity 23.2%; Pred. No. 1.7e-10;
Matches 161; Conservative 59; Mismatches 192; Indels 282; Gaps 29;

OY 100 TOAPATTOATPTTOAPTTOAPTTOA---PTTTOAPTTOAPTTOAPTTOAPT 156
Db 410 TNAATTTTAKYIFSKAPESTTSTLTGTFADPTTTGLPSTHVPNTLTAPAST-GPTV 468
OY 157 TTOAPTTOAPTTOAPTTOAPT-----TQAP--TTTOAPTTOAP--TTTQOAT 201
Db 469 STADVTSPAGTSGASPTVPSBPNDNGTESKAPMTSTSEVTTPTNATSPPTAVT 528
OY 202 TPAATTPAAT--TPAATTPAATTPAATP--GYAPATSAPVWPICELLPNGCPADFIDH 257
Db 529 TP--TPNATSPPTAVTP--TPNATSPILGKTSPTSATTP-----TN----- 569
OY 258 LILPHDKYCLFYQCSNGTYFEORCPBGILFYFNRYQCDSPANVECGELISPAAPVTEGN 317
Db 569 -----ATSPTL-----GKTSPTSAPV--- 585
OY 318 EDEDIDIGLLDNGCPANFEIDMLPHGNRCDKYQCVAHNLNVERGAGATHRSFELQOC 377
Db 585 -----TPPN----- 590
OY 378 DHELVGCTLPGESESEVDEDACTGMYCPTPEIEMELPNCGPADFSIDHLLPHESDC 437
Db 590 -----ATSP-----TLGKTSPTSAPV--TPPNAT----- 613
OY 438 GQYLOCVHGGTIFAPCGNLFHSPATDSCSPYTAGCQVECDSDNOCSTTAAPTAAPTA 497
Db 613 -----GPTVETSP-----QANATNTLTGTSPTPVTS 641
```

```
OY 498 APPTAAPTAA-----PTAAPSTVVPAPPTAAPTAVPPTTAIPT----- 536
Db 642 QPKNATSAVYTGQHNITSSSTSMSLRPPSSNPETLSPSTSDNSTSMPLLTSHAPGGEN 701
OY 536 ---PAPTAAPTAAPTAPBSPTTVTPPTAAPTAAPTAAPTAVP-EIDITVTSAP-----TA 586
Db 702 ITQVTPASISTHHVSTSSPE-PRPGTTSQASGCGNSTSTKPGEVAVVTGTTPQONATSPQ 760
OY 587 APPTAAPTAAPTAAPTAVELPT-----YTSPPT---AAPTTAAPNTTVY 632
Db 761 APGQKTAAPTAVPTVSTGKANSTTGKHTTGAGRTSTETPTDYGDSSTPRPRYNATTVL 820
OY 633 PPTAAPPTAAPTAPNTVTAAPTAAPTTAAPTAAPTVTVPTAAPTAAPTVAHAPNTTAA 692
Db 821 PPS---TSSKLRPRMTFTSPV---TTA---QATVPPTSQPRFSNLSMLVLQNASLA 870

OY 693 PVT-----TTSAPATPEDDDID 710
Db 871 VTLTLLLVADCAFRRLNSTSHYTPPYDDAE 904

RESULT 11
R07671
ID R07671 standard; protein; 278 AA.
AC R07671;
DE 22-FEB-1991 (first entry)
DE Intestinal mucin deduced from clone SMUC 41.
KM Epithelial cancer; cystic fibrosis; ulcerative colitis;
KW glycosyl transferase; familial polyposis coli;
OS Homo sapiens.
FH Key Location/Qualifiers
FT region 6..28
FT /label= repeat 1
FT region 29..52
FT /label= repeat 2
FT region 53..75
FT /label= repeat 3
FT region 76..98
FT /label= repeat 4
FT region 99..121
FT /label= repeat 5
FT modified_site 265..267
FT /label= N-glycosylation site.
PN MO9012892-A.
PD 01-NOV-1990.
PE 25-JUL-1989; U03206.
PR 14-APR-1989; US-338710.
PA (REGC) University of California.
PI Kim YS; Ginn JR.
DR WPI: 90-348495/46.
DR N-PSDB: Q06484.
DR Human intestinal mucin DNA, polypeptides and antibodies - used in
PT diagnosis and treatment of e.g. cancers, cystic fibrosis and
PT colitis.
PS Disclosure; Fig 4B; 45pp; English.
CC The sequence was deduced from a clone isolated from a cDNA library
CC using antibody probes. The DNA sequence allows determination of the
CC primary amino acid structure, difficult to characterise
CC biochemically because the protein backbone is so heavily glycosyl-
CC ated and conditions required to remove the carbohydrate tend to
CC result in breakage of the protein. Prods. of the DNA can be used
CC to produce Abs useful for detecting hnm and determining glycosyl-
CC action patterns or glycosyltransferase activity. They can be used
CC for the early detection and differential diagnosis of cancers esp.
CC epithelial cancer, and also in diseases with altered intestinal
CC mucin prodn. such as cystic fibrosis and colitis.
CC See also R07670-76.
SQ Sequence 278 AA;

Query Match 7.1%; Score 313.5; DB 1; length 278;
```

[illegible]

RESULT	12
ID	W64359
AC	W64359 standard; Protein; 267 AA.
DT	09-NOV-1998 (first entry)
DE	Mycobacterium tuberculosis antigen Tbh-29.
KM	Tuberculosis; infection; diagnosis; antigen; Tbh-29.
OS	Mycobacterium tuberculosis strain H37Rv.
PN	MO9816645-A2.
PD	23-APR-1998.
PF	07-OCT-1997; U18214.
PR	13-MAR-1997; US-818111.
PA	11-OCT-1996; US-729622.
PI	(CORI-) CORIXA CORP.
PI	Compos-Neto A, Dillon DC, Houghton R, Lodes MJ,
PI	Reed SG, Skeiky YAM, Twardzik DR, Vedvyick TS;
DR	WPI; 98-251282/22.
PT	New-PDSB: V44403.
PT	NF isolated Mycobacterium tuberculosis polypeptides and DNA - used
PT	to develop products for the detection of M. tuberculosis infection
PT	and diagnosis of tuberculosis
PS	Example 3; page 157-158; 250pp; English.
CC	This polypeptide comprises an antigenic portion of Mycobacterium
CC	tuberculosis antigen Tbh-29. A DNA molecule (see V44403) encoding
CC	the polypeptide was isolated from a M. tuberculosis strain H37Rv
CC	genomic library. The invention relates to compositions and methods
CC	for diagnosing tuberculosis. It provides polypeptides (see
CC	W64291-W64379) comprising an antigenic portion of a soluble M.
CC	tuberculosis antigen, or an immunogenic portion of an M.
CC	tuberculosis antigen, as well as DNA sequences encoding such
CC	polypeptides, recombinant expression vectors and transformed or
CC	transfected host cells. Also claimed are methods and diagnostic
CC	kits for detecting M. tuberculosis infection in a patient using
CC	these polypeptides, antibodies or oligonucleotide probes and
CC	primers, for the diagnosis of tuberculosis.
SQ	Sequence 267 AA:
Query Match	7.0%; Score 309.5; DB 1; Length 267;
Best Local Similarity	34.8%; Pred. No. 3,5e-10;
Matches	86; Conservative 11; Mismatches 107; Indels 43; Gaps
Dt	487 STAAPTAAPTAAPTAAPTAAPTAASTVVP---PATPPATAAVFP-----530A
Dt	: : : :         : : : :
Dt	12 ASLAVAVAVTIRPTAA--SKVEGHQNQAOPCKFRLPLTTQQQAAPVPPDDPTAGRGG 69
Oy	530 -----TTAIPPADPAAPTAAPTAAPESPITTVTPPTAAPTAAPTTA-----VPEIPT 578A
Dt	: : : :   : : : :   : : : :   : : : :
Dt	70 TIPAVONNVPRPGISPGVGCPASPADEAPAVPCVDPAPVIPPIIIFPGMGOMCMT 129A
Oy	579 TVTSAPTAAPTAAPTAAPTAAPTAAVEIPT-----VTSPTTAAPTAAPATVTYVP 633A
Dt	: : : :   : : : :   : : : :   : : : :
Dt	130 IPTAPPTPVTTSATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTPTV-AP 168A

QY	634	PTAAPTAAPBNTTVAAPPTAAPTAAAPBNTTVAAPPTAAAPPTAAAPPTAAAPBNTTVAAP	633
Db	189	TTVAPTTVAAP---TTVAAPTVAAPATATP---TTVAAPQPTQPTQPTQPTQPTQPTQPTVAAP	241
QY	694	VTTSAP	700
Db	242	QTVAAP	248

RESULT 13  
W81726  
ID W81726 standard; Protein; 267 AA.  
AC W81726;  
DT 27-JAN-1999 (first entry)  
DE M. tuberculosis immunogenic polypeptide TBH-29.  
KW Tuberculosis; immunogenic; soluble; antigen; protective immunity; TB;  
KW vaccine; pharmaceutical; infection; diagnosis.  
OS Mycobacterium tuberculosis.  
PN W09816646-A2.  
PD 23-APR-1998.  
PF 07-OCT-1997; U18293.  
PR 13-MAR-1997; US-818112.  
PR 11-OCT-1996; US-730510.  
PA (CORI-) CORIXA CORP.  
PI Campos-Neto A, Dillon DC, Houghton R, Loëes MJ,  
PI Reed SG, Skeiky YAM, Twardzik DR, Vedvick TS;  
DR WPI: 98-261042/23.  
DR N-PSDB: V64512.  
PT Immunogenic Mycobacterium tuberculosis polypeptide(s) and DNA - used  
PT to develop products for the detection of M. tuberculosis infection  
PI and for diagnosis, treatment and prevention of tuberculosis  
PS Example 3b, Page 149-150; 230pp: English.  
CC This sequence represents an immunogenic portion of a soluble  
CC Mycobacterium tuberculosis (MT) antigen which can be used in a method  
CC for inducing protective immunity against tuberculosis (TB).  
CC This sequence can be formulated into vaccines and/or pharmaceutical  
CC compositions for immunising against M. tuberculosis infection or may  
CC be used for the diagnosis of tuberculosis.  
SQ Sequence 267 AA;

[illegible]





**THIS PAGE BLANK (USPTO)**



```

Db      569  -----ATSPTL-----GKSPSTSAVT---565
QY      318  EDEDIDIDDLNCGCPANFEDIMLLPHGNRCDKITYQCYHGNLVERRCAGATHSFELQOC 3777
Db      585  -----TPPN-----590
QY      378  DHIELVGCTLPGESEEEVDVEDACTGWCYCTPEPLEMEPLPNCGRAPFSIDHLLPHESDC 437
Db      590  -----ATSP-----TLGKTSPTSVAVT-TPPDNAT-----613
QY      438  GQYLQCVHGQTIARPCPGNLHFSPTATOCESPVTAGCOVEPCDSNOCSTAAFTAFTA 497
Db      613  -----GPIVGERSP-----QANTNHTLGGTSPTPVYTS 6411
QY      498  APTAAPTAA-----PTAAPSTVVPAPPAFTAAVPPTTAIPT-----5366
Db      642  QPKNNTSAVTTGQHNHTSSSTSMSLRPSNPETLSPTSDNSTSHMPLTISAHPTEGEN 7010
QY      536  ---PAPTAAPTAAPTAADESPTVTVPTTAAPTAAPTANP-EIPIVYSAP-----TA 5066
Db      702  ITQVTPASISTHNVSTSSPE-PRGRTISQASGPGNSSSTSPGWNVYTKGPONNATSPQ 7606
QY      587  APTAAPTAAPTAAPTAAPEIPIPT-----TVTSPPR---AAPTTAAPAPNTTVV 6322
Db      761  APSGCKTAVPVYTSIGGANSTGCKHNTGCGARTSTEPTTDYGDSTTPPRFNATTYL 8200
QY      633  PPTAAPTTAAPAPNTTVTAPPTAAPTTAAPAPNTTVTVPPTAAPTAAPTVAHAAPNTTAA 6922
Db      821  PPS---TSSKLRPRTFTSPPV---TTA---QATVPVPPTSGPRFSNLSMLVLQMASLA 8700
QY      693  PVT-----TTSAPATPEDDDD 710
Db      871  VTLTLLLVMAACAFAFRNLSSTSHYTTTPPYDAE 904

RESULT      2
US-08-287-001a-2
: Sequence 2, Application US/08287001a
: Patent No. 5622861
: GENERAL INFORMATION:
: APPLICANT: KAPLAN, GERARDO
: APPLICANT: FEINSTONE, STEPHEN M.
: TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
: TITLE OF INVENTION: OF USE
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C.
: STREET: Suite 1200, The Candler Bldg, 127 Peachtree
: STREET: Street
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/287,001a
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Spratt, Gwedolyn D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 1414,621
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404/688-0770
: TELEFAX: 404/688-9880
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 451 amino acids
: TYPE: amino acid
:

```

```

:          TOPOLOGY: linear
:          MOLECULE TYPE: protein
:          US-08-287-001A-2

Query Match          7.0%; Score 308.5; DB 1; Length 451;
Best Local Similarity 41.6%; Pred. No. 1.4e-13;
Matches 87; Conservative 4; Mismatches 53; Indels 65; Gaps 10;

OY 94 GPPAET-----TQAPATQAPTTQAPTTQAPTTQAPTTQAPTTQAPTTT 139
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 125 GPPRVTEIVRVRSTVPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTT 183
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 140 QAPTTQAPTTQAPTT---TQAPTTTQAPTTT---QAPTTQAPTTQAPTTQ 191
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 184 TPTTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPTTTLPT 242
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 192 APTTTQAATPAATP-----PAAT-PAATPAAT-----PAAT 226
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 243 LPTTLPTMTLPTTTLPTTMTLPTTTLPTTTLPTTTLPTTMTVSTVPPPLPMQNHVPAT 302
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
OY 227 TPGVP-----APTSAPVW 239
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
DB 303 SPSSPQPAETHPVTLTGATRTQPTSSPLX 331
      |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 3
PCT-US95-09941-2
: Sequence 2, Application PC/TUS9509941
: GENERAL INFORMATION:
:   APPLICANT:
:   TITLE OF INVENTION: HEPATITIS A VIRUS RECEPTOR AND METHODS
:   TITLE OF INVENTION: OF USE
:   NUMBER OF SEQUENCES: 5
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: NEEDLE & ROSENBERG, P.C.
:   STREET: Suite 1200, The Candler Bldg, 127 Peachtree
:   STREET: Street
:   CITY: Atlanta
:   STATE: Georgia
:   COUNTRY: USA
:   ZIP: 30303
:   COMPUTER READABLE FORM:
:   MEDIUM TYPE: Floppy disk
:   COMPUTER: IBM PC compatible
:   OPERATING SYSTEM: PC-DOS/MS-DOS
:   SOFTWARE: Patentln Release #1.0, Version #1.25
:   CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: PCT/US95/09941
:   FILING DATE:
:   CLASSIFICATION:
:   PRIOR APPLICATION DATA:
:   APPLICATION NUMBER: 08/287,001
:   FILING DATE: 5 AUG 1994
:   ATTORNEY/AGENT INFORMATION:
:   NAME: Spiralt, Gwedolyn D.
:   REGISTRATION NUMBER: 36,016
:   REFERENCE/DOCKET NUMBER: 1414, 621
:   TELECOMMUNICATION INFORMATION:
:   TELEPHONE: 404/688-0770
:   TELEFAX: 404/688-9880
:   INFORMATION FOR SEQ ID NO: 2:
:   SEQUENCE CHARACTERISTICS:
:   LENGTH: 451 amino acids
:   TYPE: amino acid
:   TOPOLOGY: linear
:   MOLECULE TYPE: protein
:   PCT-US95-09941-2

Query Match          7.0%; Score 308.5; DB 3; Length 451;
Best Local Similarity 41.6%; Pred. No. 1.4e-13;
Matches 87; Conservative 4; Mismatches 53; Indels 65; Gaps 10;

```

OY 94 GPPAET-----TOAPATTQAPTTTQAPTTTQAAPT---TQAAPT 139  
 125 GPRVATPIVIRVSTVTTTTLPTTTTLPTTTT-LPTTTTLPMTTTTLPTTT 183  
 OY 140 QAATTTQAPTTTQAAPT---TQAPTTTQAPTTT---QAPTTTQAAPT 191  
 184 TVTTLPTTTTLPTTTTLPTTTTLPTTT-LPTTTTLPTTTTLPTTTTLPTTT 242  
 OY 192 APTTTQAATTPAAT-----PAAT--PAATTPAAT-----PAAT 226  
 243 LPTTTLPTTTTLPTTTTLPTTTTLPTTTTLPTTTVSTFVPTPLPMONHEPAT 302  
 OY 227 TPGVP-----APTSAPVW 239  
 303 SPSSQPAETHPVTLTGARTQPTSSPLX 331  
 Db

RESULT 4  
 US-08-276-967-2  
 : Sequence 2, Application US/08276967  
 : Patent No. 5851817  
 : GENERAL INFORMATION:  
 : APPLICANT: Hardy, Daniel M.  
 : APPLICANT: Garbers, David L.  
 : TITLE OF INVENTION: Species-Specific Egg-Binding Proteins of  
 : TITLE OF INVENTION: Sperm  
 : NUMBER OF SEQUENCES: 8  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Arnold, White & Durkee  
 : STREET: P. O. Box 4433  
 : CITY: Houston  
 : STATE: Texas  
 : COUNTRY: USA  
 : ZIP: 77210-4433  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/276,967  
 : FILING DATE: Submitted Herewith  
 : CLASSIFICATION: 530  
 : ATTORNEY/AGENT INFORMATION:  
 : NAME: Kitchell, Barbara S.  
 : REGISTRATION NUMBER: 33,928  
 : REFERENCE/DOCKET NUMBER: UTSD:418\KIT  
 : TELECOMMUNICATION INFORMATION:  
 : TELEPHONE: 713-787-1400  
 : TELEFAX: 713-789-2679  
 : TELEX: 79-0924  
 : INFORMATION FOR SEQ ID NO: 2:  
 : SEQUENCE CHARACTERISTICS:  
 : LENGTH: 2476 amino acids  
 : TYPE: amino acid  
 : STRANDEDNESS: single  
 : TOPOLOGY: linear  
 : MOLECULE TYPE: protein  
 : US-08-276-967-2

Query Match 6.5%; Score 283.5; DB 2; Length 2476;  
 Best Local Similarity 20.9%; Pred. No. 4.2e-11;  
 Matches 152; Conservative 42; Mismatches 212; Indels 323; Gaps 31;  
 OY 90 CULGPPAETTPAAT-ATQAAPT-----TQAPTTTQAPTTTQAAPT--T 138  
 310 CRGPGSTSVSTKPAVTEKPTVSEIYITPEKPMVHMEKPIVHNEKPIVPEKPTIPT 369  
 OY 139 TQAPTTTQAPTTTQAPTTTQAPTTTQAAPT--TQAAPT--TQAAPT--TQAAPT 195  
 370 EKSTVPTKPTVKEPTELPPEGPTVAERPTPEGPAVPKGPVLTLEMTSTHTEKSTV 429  
 Db

OY 196 ITQAATTPAATPAAT--TPAATTPAATTPAATPGVAPATSAPVPPICELLPNGCPADF 254  
 430 HTEKPLIPGKSTIPTEKPMVPTKRTTPTERT----- 463  
 OY 255 DIHLILPHDKYCNLFYQCSNGTYTEQRCPEGLFYNDYORCSPANVECDGEISAPAVT 314  
 463 ---TIPAEK----- 469  
 OY 315 EGNEDDIDIGDLNDGCPANFEIDWLPHGNRCDKYQCVHGNLVERRCGATHSFEL 374  
 469 ----- 469  
 OY 375 QQCDEHIELVCTLPGGESEVDVEDACTGWYCPTPEIMEPLPNGCPADFSIDLHPHE 434  
 469 -----PTVPIEKPMV----- 480  
 OY 435 SDCGYLQCVHGQTARPCPGNLHSPATQSCSPVTAGCQVFECDSDNQCSTAAPTAA 494  
 480 -----TERTTIPT-----ERTTIPT 496  
 OY 495 PTAP 547  
 497 PT-VPTKEKL-VTERK-IVP-----TEKPIVPTKEHTIPEKLTVLTERTTPTERT 546  
 OY 548 TAAPESPTTVTPPTAPAPAPAP--TAVPEIP---TIVTSAPAPAPAPAPAPAPAP 602  
 547 TITPEKPTVPTKEP-SVPEKPTVPEEPTIPEKLTVTERTTPTTKR-TITPTIRITT 604  
 OY 603 AVEIPTVTSPTAPAP--TAAPAPNTVTPPTAPAPAPAPAPAPAPAPAPAPAPAP 659  
 605 ---PTIRTTPTERTTPTIRTTPTERTTPTIRTTPTIRTTPTIRTTPTIRTTPTIRTT 662  
 OY 660 AAPAPNTVTPPTAPAPAP--PTVAHAP-----NTTAPVTTTSAPATPEDDIDP- 713  
 663 LVFTQPAAVMSTSATVTPRTTASCPNNAHFERCACPV---SCQSPTPNCELECKEG 719  
 OY 713 -----LPNDPINP-CVEECANVLPNAHDCDKY--- 739  
 Db 720 CVCDDPGLTSSGSHCVNASSCCDFYNDNYTKLGTDWSPNCTEHCRCPSRMECQIFKCG 779  
 OY 739 --WVCDGNN 745  
 Db 780 THTVCOLKN 788  
 Db

RESULT 5  
 US-08-217-327-4  
 : Sequence 4, Application US/08217327  
 : Patent No. 5474925  
 : GENERAL INFORMATION:  
 : APPLICANT: John, Maliyakal E  
 : APPLICANT: Barton, Kenneth A  
 : TITLE OF INVENTION: Immobilized Proteins in Cotton Fiber  
 : NUMBER OF SEQUENCES: 16  
 : CORRESPONDENCE ADDRESS:  
 : ADDRESSEE: Quarles and Brady  
 : STREET: P.O. Box 2113  
 : CITY: Madison  
 : STATE: WI  
 : COUNTRY: USA  
 : ZIP: 53701-2113  
 : COMPUTER READABLE FORM:  
 : MEDIUM TYPE: Floppy disk  
 : COMPUTER: IBM PC compatible  
 : OPERATING SYSTEM: PC-DOS/MS-DOS  
 : SOFTWARE: Patent Release #1.0, Version #1.25  
 : CURRENT APPLICATION DATA:  
 : APPLICATION NUMBER: US/08/217,327  
 : FILING DATE:  
 : CLASSIFICATION: 435  
 : PRIOR APPLICATION DATA:  
 : APPLICATION NUMBER: US 07/812,233

```

1      FILING DATE: 19-DEC-1991
2      ATTORNEY/AGENT INFORMATION:
3      NAME: Seay, Nicholas J
4      REGISTRATION NUMBER: 27,386
5      REFERENCE/DOCKET NUMBER: 1122990831
6      TELECOMMUNICATION INFORMATION:
7      TELEPHONE: 608-251-5000
8      TELEFAX: 608-251-9166
9      INFORMATION FOR SEQ ID NO: 4:
10     SEQUENCE CHARACTERISTICS:
11     LENGTH: 214 amino acids
12     TYPE: amino acid
13     TOPOLOGY: linear
14     MOLECULE TYPE: protein
15     US-08-217-327-4
16
17 Query Match      6.3%; Score 276.5; DB 1; Length 214;
18 Best Local Similarity 39.2%; Pred. No.7.9e-12;
19 Matches 80; Conservative 12; Mismatches 67; Indels 45; Gaps 11;
20
21 Oy 502 APTAAPATPSTVPPATPPATAPATVPPPTTAIPATPAPATPAAATPAAESPTTVVPP 561
22      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
23 Db 25 APSNPSTPTPTPTPPASNP-----PPTQAP-DTPRTP-----PVSPTPP 66
24
25 Oy 562 TAAPTAATTVAPELPIVTSAPTAAPTAAPTAATTAATTVAPELPIVTSPPTAAPT 621
26      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
27 Db 67 TSSP--PVTASP--PVSTPPSSPPPATP--PPASPPPATP---PPASPPPATPP 115
28
29 Oy 622 AAPAPNTVTVPPTAAPTAAPAPNTVTAAPTAAPT--AAPAPNTVTVPPTAAPTA 679
30      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
31 Db 116 ASPPAT-----PPATPPPATPPPATPPAPPLASPPATVPALSPVQTPLTSP-----A 165
32
33 Oy 680 PPTVAHAPNTTAAPVTTGSAPATT 703
34      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
35 Db 166 PTEAPAPTLGNA---TPGPAGT 185
36
37 RESULT 6
38 US-08-479-537A-2
39 Sequence 2, Application US/08479537A
40 Patent No. 5861381
41 GENERAL INFORMATION:
42 APPLICANT: CHABON, Pierre
43 APPLICANT: KIENE, Marie-Paule
44 APPLICANT: LAIHE, Richard
45 APPLICANT: HAREUVENI, Mara
46 TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE
47 TREATMENT OR PREVENTION OF A MALIGNANT TUMOR
48 NUMBER OF SEQUENCES: 5
49 CORRESPONDENCE ADDRESS:
50 ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS, L.L.P.
51 STREET: P.O. Box 1404
52 CITY: Alexandria
53 STATE: Virginia
54 COUNTRY: United States
55 ZIP: 22313-1404
56 COMPUTER READABLE FORM:
57 MEDIUM TYPE: floppy disk
58 COMPUTER: IBM PC compatible
59 OPERATING SYSTEM: PC-DOS/MS-DOS
60 SOFTWARE: PatentIn Release #1.0, Version #1.30
61 CURRENT APPLICATION DATA:
62 APPLICATION NUMBER: US/08/479,537A
63 FILING DATE: 07-JUN-1995
64 CLASSIFICATION: 514
65 PRIOR APPLICATION DATA:
66 APPLICATION NUMBER: FR 90/13101
67 FILING DATE: 23-OCT-1990
68 PRIOR APPLICATION DATA:
69 APPLICATION NUMBER: WO PCT/FR91/00835
70 FILING DATE: 23-OCT-1991
71 PRIOR APPLICATION DATA:

```

```

1 APPLICATION NUMBER: US 08/039,320
2 FILING DATE: 04-APR-1993
3 PRIOR APPLICATION DATA:
4 APPLICATION NUMBER: US 08/403,576
5 FILING DATE: 14-MAR-1995
6 ATTORNEY/AGENT INFORMATION:
7 NAME: Teskin, Robin L.
8 REGISTRATION NUMBER: 35,030
9 REFERENCE/DOCKET NUMBER: 017753-025
10 TELECOMMUNICATION INFORMATION:
11 TELEPHONE: (703) 836-6620
12 TELEFAX: (703) 836-6620
13 INFORMATION FOR SEQ ID NO: 2:
14 SEQUENCE CHARACTERISTICS:
15 LENGTH: 2035 amino acids
16 TYPE: amino acid
17 STRANDEDNESS: single
18 TOPOLOGY: linear
19 MOLECULE TYPE: peptide
20 FEATURE:
21 NAME/KEY: Peptide
22 LOCATION: 128..1899
23 OTHER INFORMATION: /note="The amino acids spanning
24 OTHER INFORMATION: 128 to 1899 constitute a repeated region wherein the repeat
25 OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
26 OTHER INFORMATION: repeats varies from 1 to 40."
27 FEATURE:
28 NAME/KEY: Peptide
29 LOCATION: 134
30 OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
31 OTHER INFORMATION: Xaa Xaa which is the codon for Pro or Ala wherein Pro = CC
32 OTHER INFORMATION: CCC, CCA, or CCG; and Ala = GCT, GCC, GCA, or GCG."
33 FEATURE:
34 NAME/KEY: Peptide
35 LOCATION: 144
36 OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
37 OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
38 OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
39 FEATURE:
40 NAME/KEY: Peptide
41 LOCATION: 147
42 OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
43 OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
44 OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
45 FEATURE:
46 NAME/KEY: Peptide
47 LOCATION: 1..21
48 OTHER INFORMATION: /note="Amino acids 1 to 21 are a
49 OTHER INFORMATION: 21 amino acid precursor sequence."
50 OTHER INFORMATION:
51 US-08-4/9-537A-2
52
53 Query Match 6.3% Score 275.5; DB 2; Length 2035;
54 Best Local Similarity 23.5%; Pred. No. 1.1e-10;
55 Matches 166; Conservative 45; Mismatches 291; Indels 203; Gaps 35;
56
57 QY 93 GPPAETTQATTAATTQAPTTQAPTT-----TTQAPTTTQAPTTTQAPTTTQAPTTT 145
58 || : ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
59 Db 66 PGSSSTTQGGDVLAPRTEPASAASATWGDVTSVPTTRALSTTPRAIDVTSAPDNK 125
60
61 QY 146 QAPTTQAPTT--TTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP-----TTTQ 198
62 || : ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
63 Db 126 PAPGST-APXAHGVTSAADXRPXPGSTAPXAHGVTSAADXRPXPGST-APXAHGVTSA 183
64
65 QY 199 AATTPAATTTPA-----TTPAATTTPA-----TTPAATTTPGVPAPTSAPV 238
66 || : ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
67 Db 184 XRPXPGSTAPXAHGVTSAADXRPXPGSTAPXAHGVTSAADXRPXPGSTAPXAHGVTSA 243
68
69 QY 239 WPPICELLNCCPADEFDILLIPHDKYCNLFYQCSNGYTFQRCEDEGLFNPYVQRODSP 298
70 || : ||| ||| ||| : ||| ||| ||| ||| ||| ||| ||| ||| |||
71 Db 244 XRPX-----PGSTAPX-----AHGVT-----SAP 262
72
73 QY 299 ANVEDGEISAPPTGNEDEDIDGLDNGGCPANFEIDWLLPHGNRCDKYQCVHGN 358

```

```

Db 263 DXRPGSTAPX-----AHG- 278
QY 359 LVERRCGAGTHSFELQOCDHIELVGCITLPGSESEVVDVEDACTGWYCP-----TEPIE 413
Db 278 -----VTSAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPX 323
QY 414 WEPLPNC-CPADFSIDHLLPHESDCGQYLOCVHGGTIA---RPCGNLHESPAITQSCSP 469
Db 324 XRPXPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGST 373
QY 470 VTGCGCYFECDSDNOCTSTAPTA-APTAPTAAP---TAAPTAAPSTVVPATP-PATA 524
Db 374 XAHG--VTSAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGST 431
QY 525 AP-VPTTAIP--TPAP-TAPTAAPTTAPAESPTVTPPTAPPTAPTAAPTAAPTAAPTAAPTAAPTAAP 579
Db 432 APXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSA 488
QY 580 VTSAP-----TAAPTAAP--TAAPTAAPTTAPVEI-PTVTSPTAPPTAA-----PAP 626
Db 489 GSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHG 548
QY 627 NT-----VTVPPTAPPTAPTAAPNT-TTAPPTAPPTAPTAAPNT-TTVPPTAP-- 677
Db 549 GSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHG 608
QY 677 -TAAPNTTAAPVT-----TSAPATPEDDIDP 711
Db 609 GSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAP 653

```

## RESULT

US-08-479-537A-5

Sequence 5, Application US/08479537A

Patent No. 5861381

GENERAL INFORMATION:

APPLICANT: CHAMON, Pierre

APPLICANT: KIENY, Marie-Paule

APPLICANT: LATHE, Richard

APPLICANT: HAREVIENT, Mara

TITLE OF INVENTION: PHARMACEUTICAL COMPOSITION FOR THE

TITLE OF INVENTION: TREATMENT OR PREVENTION OF A MALIGNANT TUMOR

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: BURNS, DOANE, SWECKER &amp; MATHIS, L.L.P.

STREET: P.O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,537A

FILING DATE: 07-JUN-1995

CLASSIFICATION: 514

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: FR 90/13101

FILING DATE: 23-OCT-1990

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: NO PCT/FR91/00835

FILING DATE: 23-OCT-1991

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/039,320

FILING DATE: 04-APR-1993

PRIORITY APPLICATION DATA:

APPLICATION NUMBER: US 08/403,576

FILING DATE: 14-MAR-1995

ATTORNEY/AGENT INFORMATION:

```

NAME: Teskin, Robin L.
REGISTRATION NUMBER: 35,030
REFERENCE/DOCKET NUMBER: 017753-025
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 836-6620
TELEFAX: (703) 836-2021
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1867 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
FEATURE:
NAME/KEY: Peptide
LOCATION: 128..1727
OTHER INFORMATION: /note="The amino acids spanning
OTHER INFORMATION: 128 to 1727 constitute a repeated region wherein the repea
OTHER INFORMATION: 20 amino acids, 17 of which are fixed. The number of such
OTHER INFORMATION: repeats varies from 1 to 40."
FEATURE:
NAME/KEY: Peptide
LOCATION: 134
OTHER INFORMATION: /note="Amino acid 134 is X1 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 144
OTHER INFORMATION: /note="Amino acid 144 is Y = Xaa
OTHER INFORMATION: which is the codon for Thr or Asn wherein Thr = ACT, ACC,
OTHER INFORMATION: or ACG; and Asn = AAT or AAC."
FEATURE:
NAME/KEY: Peptide
LOCATION: 147
OTHER INFORMATION: /note="Amino acid 147 is X2 = Xaa
OTHER INFORMATION: which is the codon for Pro or Ala wherein Pro = CCT, CCC,
OTHER INFORMATION: or CCG; and Ala = GCT, GCC, GCA, or GCG."
FEATURE:
NAME/KEY: Peptide
LOCATION: 1..21
OTHER INFORMATION: /note="Amino acids 1 to 21 are a
OTHER INFORMATION: 21 amino acid precursor sequence."
US-08-479-537A-5

Query Match 6.3%; Score 275.5; DB 2; Length 1867;
Best Local Similarity 23.5%; Pred. No. 1e-10;
Matches 166; Conservative 45; Mismatches 291; Indels 203; Gaps 35;

```

```

QY 93 RGPAPTQAPATTAAPTQAPT-----TTQAPTTTQAPTQAPTQAPTQAPT 145
Db 66 PGSGSSITGQDVTLPATPASPAGSAATWGQDVTSPVTRPALGSTITPPAHDTASAPDK 125
QY 146 QAPTTQAPT--TTOAPTQAPTQAPTQAPTQAPTQAPTQAPTQAPTQAPT 198
Db 126 PARGST-APXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGST-APXAHGVTSAAD 183
QY 199 AATTPAATTPAA-----TTAATTPAA-----TTAATTPGVPAPTSAV 238
Db 184 XRPXPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAD 243
QY 239 WPICELPNCPADFDHLLPHDXKYNLFYCCSNGYFEQNCPPGGLVFNPPVQCDSP 298
Db 244 XRPX-----PGSTAPX-----AHGVT-----SAP 262
QY 299 ANECDGEISPAPTYEGNEDDIDIGLLDNGCPANFEIDMLPHGNRCCKXYQCVMGN 358
Db 263 DXRPGSTAPX-----AHG- 278
QY 359 LVERRCGAGTHSFELQOCDHIELVGCITLPGSESEVVDVEDACTGWYCP-----TEPIE 413
Db 278 -----VTSAPDXRPGSTAPXAHGVTSAAPDXRPGSTAPXAHGVTSAAD 323

```













```

QY 567 AAPTAVPEIPTVTSAPTAAPTAAPTAAPTAAPTAAVE--IPTVTS----PTAAPT 620
    : ||| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 184 SKPTTQRFLLPPEVTKAATIRATLET-----PTTISIPETSVLFEVTSRLVPSSTAKP 237
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 621 TAAPANTTVTPPTAAPTAAAPANTTVTAAPTAAPTAAAPANTTVTPPTAAPTAA 680
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 238 RSLPKPSTSTAEPTKSTA--LPSSPTLLPPTAEP-OVEPELTTVLDSLEVPPTSSG 293
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 681 P-----TVAHAPNTTAAPVYTTTSAP 700
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 294 PSGDFEIQEEETTRPELGNVAVVTPPAAP 325
    : | | | | | | | | | | | | | | | | | | | | | | | | | | | |

```

Search completed: July 18, 1999, 07:57:25  
 Job time: 446 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 03:46:46 ; Search time 42.46 Seconds

(without alignments)  
923.231 Million cell updates/sec

Title: US-09-294-663-4

Perfect score: 4395  
Sequence: 1 MIKILLFLTAGLVARREV.....VFIPWKLDEIDRALNFEL 807

Scoring table: BLOSUM62

Searched: 396714 seqs, 48575446 residues

Database : Pending\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/paa/PCFUS9.COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/paa/US07.COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/paa/US08.COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/paa/US081.COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/paa/US082.COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/paa/US083.COMB.pep.\*
- 7: /cgn2\_6/ptodata/1/paa/US084.COMB.pep.\*
- 8: /cgn2\_6/ptodata/1/paa/US084B.COMB.pep.\*
- 9: /cgn2\_6/ptodata/1/paa/US085.COMB.pep.\*
- 10: /cgn2\_6/ptodata/1/paa/US086.COMB.pep.\*
- 11: /cgn2\_6/ptodata/1/paa/US087.COMB.pep.\*
- 12: /cgn2\_6/ptodata/1/paa/US088.COMB.pep.\*
- 13: /cgn2\_6/ptodata/1/paa/US089.COMB.pep.\*
- 14: /cgn2\_6/ptodata/1/paa/US090.COMB.pep.\*
- 15: /cgn2\_6/ptodata/1/paa/US091.COMB.pep.\*
- 16: /cgn2\_6/ptodata/1/paa/US092.COMB.pep.\*
- 17: /cgn2\_6/ptodata/1/paa/US093.COMB.pep.\*
- 18: /cgn2\_6/ptodata/1/paa/US60.COMB.pep.\*
- 19: /cgn2\_6/ptodata/1/paa/PCF\_NEM.COMB.pep.\*
- 20: /cgn2\_6/ptodata/1/paa/PCF\_NEM.COMB.pep.\*
- 21: /cgn2\_6/ptodata/1/paa/US08\_NEM.COMB.pep.\*
- 22: /cgn2\_6/ptodata/1/paa/US09\_NEM.COMB.pep.\*
- 23: /cgn2\_6/ptodata/1/paa/US60\_NEM.COMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	4387	99.8	807	16	US-09-294-663-4	Sequence 4, Appl1
2	4352	99.0	805	15	US-09-103-429A-4	Sequence 4, Appl1
3	4266.5	97.1	788	16	US-09-294-663-3	Sequence 3, Appl1
4	4231.5	96.3	786	15	US-09-103-429A-3	Sequence 3, Appl1
5	542	12.3	1404	2	US-07-643-502C-13	Sequence 13, Appl
6	542	12.3	1205	2	US-07-643-502C-28	Sequence 28, Appl
7	542	12.3	1404	2	US-07-757-022B-2	Sequence 2, Appl1
8	542	12.3	941	2	US-07-757-022B-14	Sequence 14, Appl
9	542	12.3	1361	2	US-07-757-022B-40	Sequence 40, Appl
10	542	12.3	1311	2	US-07-757-022B-42	Sequence 42, Appl
11	542	12.3	1270	2	US-07-757-022B-44	Sequence 44, Appl
12	542	12.3	1320	2	US-07-757-022B-46	Sequence 46, Appl
13	542	12.3	1354	2	US-07-757-022B-48	Sequence 48, Appl
14	542	12.3	1314	2	US-07-757-022B-50	Sequence 50, Appl
15	542	12.3	1363	2	US-07-757-022B-52	Sequence 52, Appl
16	542	12.3	1049	2	US-07-757-022B-56	Sequence 56, Appl
17	542	12.3	1320	2	US-07-757-022B-60	Sequence 60, Appl
18	542	12.3	1404	2	US-07-757-022B-62	Sequence 62, Appl
19	542	12.3	1038	2	US-07-757-022B-74	Sequence 74, Appl
20	542	12.3	1022	2	US-07-757-022B-84	Sequence 84, Appl

21	542	12.3	1140	2	US-07-757-022B-104	Sequence 104, App
22	542	12.3	1313	2	US-07-757-022B-142	Sequence 142, App
23	463	10.5	1837	13	US-08-928-361-5	Sequence 5, Appl1
24	463	10.5	1837	21	US-08-928-361B-5	Sequence 5, Appl1
25	453.5	10.3	1721	1	PCR-US97-14104-5	Sequence 5, Appl1
26	453.5	10.3	1721	11	US-08-700-651A-5	Sequence 5, Appl1
27	453.5	10.3	1721	11	US-08-700-651-5	Sequence 5, Appl1
28	453.5	10.3	1721	13	US-08-928-361-6	Sequence 6, Appl1
29	453.5	10.3	1721	21	US-08-928-361B-6	Sequence 6, Appl1
30	410	9.3	1367	18	US-60-087-236-55	Sequence 55, Appl
31	372	8.5	557	16	US-09-248-796-26892	Sequence 26892, A
32	372	8.5	557	18	US-60-096-409-26892	Sequence 26892, A
33	339	7.7	827	16	US-09-248-796-17307	Sequence 17307, A
34	339	7.7	827	18	US-60-096-409-17307	Sequence 17307, A
35	335	7.6	3880	3	US-08-028-021-1	Sequence 1, Appl1
36	332.5	7.6	297	9	US-08-580-545B-6	Sequence 6, Appl1
37	332.5	7.6	297	16	US-09-262-653A-6	Sequence 6, Appl1
38	325	7.4	907	11	US-08-783-774-2	Sequence 2, Appl1
39	309.5	7.0	267	1	PCR-US99-03265-137	Sequence 137, App
40	309.5	7.0	267	12	US-08-818-111-137	Sequence 137, App
41	309.5	7.0	267	12	US-08-818-112-142	Sequence 142, App
42	309.5	7.0	267	13	US-08-928-361-8	Sequence 8, Appl1
43	309.5	7.0	267	13	US-08-942-341-137	Sequence 137, App
44	309.5	7.0	267	13	US-08-942-578-142	Sequence 142, App
45	309.5	7.0	267	13		

# ALIGNMENTS

RESULT 1  
US-09-294-663-4  
Sequence 4, Application US/09294663  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R  
APPLICANT: Wang, Ping  
TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
TITLE OF INVENTION: CDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Plinisi & Michaels, P.C.  
STREET: 118 North Tloga Street  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/294, 663  
FILING DATE: 19-APR-1999  
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 09/103,429  
FILING DATE: 24-JUN-1998  
ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39-CIP  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-2000  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 807 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHETICAL: NO

ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-294-663-4

Query Match 99.8%; Score 4387; DB 16; Length 807;  
Best Local Similarity 100.0%; Pred. No. 3.3e-273;  
Matches 807; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1 MIKLLFLTAAGLVAAAREVSDAKNPALHPHDPXPAEQXLLPXYDCTKFFYCYG 60  
1 MIKLLFLTAAGLVAAAREVSDAKNPALHPHDPXPAEQXLLPXYDCTKFFYCYG 60  
61 LKFIAPRDCAGTEFEKSAQCVAAALAGCTLPGPAAETTOAPATTOAPTTTQA 120  
61 LKFIAPRDCAGTEFEKSAQCVAAALAGCTLPGPAAETTOAPATTOAPTTTQA 120  
61 LKFIAPRDCAGTEFEKSAQCVAAALAGCTLPGPAAETTOAPATTOAPTTTQA 120  
121 APTTTOAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 180  
121 APTTTOAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 180  
121 APTTTOAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 180  
181 PTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 240  
181 PTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 240  
181 PTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 240  
241 PICELLPNGCPADFDHLLPHDKYCNLFYQCSNGYFEORCPGLYFNRYVQCDSPAN 300  
241 PICELLPNGCPADFDHLLPHDKYCNLFYQCSNGYFEORCPGLYFNRYVQCDSPAN 300  
241 PICELLPNGCPADFDHLLPHDKYCNLFYQCSNGYFEORCPGLYFNRYVQCDSPAN 300  
301 VECDEGISAPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
301 VECDEGISAPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
301 VECDEGISAPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
361 ERRCGAGTHSFELQOCDHIELVAGCTLPGGESEVVDVEDACTGMYCTEPIEMEPPLNG 420  
361 ERRCGAGTHSFELQOCDHIELVAGCTLPGGESEVVDVEDACTGMYCTEPIEMEPPLNG 420  
361 ERRCGAGTHSFELQOCDHIELVAGCTLPGGESEVVDVEDACTGMYCTEPIEMEPPLNG 420  
421 CPADFSIDHLLPHESDGOYLQCVHAGCTIARPCGNLHFSFATOSCESPVGACVFECD 480  
421 CPADFSIDHLLPHESDGOYLQCVHAGCTIARPCGNLHFSFATOSCESPVGACVFECD 480  
421 CPADFSIDHLLPHESDGOYLQCVHAGCTIARPCGNLHFSFATOSCESPVGACVFECD 480  
481 SDNOCSTAAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA 540  
481 SDNOCSTAAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA 540  
541 APTAAPTAAAPESPTVTPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
541 APTAAPTAAAPESPTVTPPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600  
601 TTAAPEIPTVTSPTAAPTAAAPNTTVTPPTAAPTAAAPNTTVTPPTAAPTAA 660  
601 TTAAPEIPTVTSPTAAPTAAAPNTTVTPPTAAPTAAAPNTTVTPPTAAPTAA 660  
661 APADNTTVTPPTAAPTAAAPNTTVTPPTAAPTAAAPNTTVTPPTAAPTAA 720  
661 APADNTTVTPPTAAPTAAAPNTTVTPPTAAPTAAAPNTTVTPPTAAPTAA 720  
721 CVEECNLLPAAHACDCKWVCDGNNQVLYVCSGLQFPTTKTCDFACNVCSNIONS 780  
721 CVEECNLLPAAHACDCKWVCDGNNQVLYVCSGLQFPTTKTCDFACNVCSNIONS 780  
781 ESEYGOVEIIPMKNLDEDIROALNFE 807  
781 ESEYGOVEIIPMKNLDEDIROALNFE 807  
781 ESEYGOVEIIPMKNLDEDIROALNFE 807

RESULT 2  
US-09-103-429A-4  
Sequence 4, Application US/09103429A  
GENERAL INFORMATION:  
APPLICANT: Granados, Robert R

APPLICANT: Wang, Ping  
TITLE OF INVENTION: A Novel Invertebrate Intestinal Mucin  
TITLE OF INVENTION: CDNA and Related Products and Methods  
NUMBER OF SEQUENCES: 4  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Brown, Pinnisi & Michaels, P.C.  
STREET: 118 North Tlloga  
CITY: Ithaca  
STATE: NY  
COUNTRY: USA  
ZIP: 14850

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/103,429A  
FILING DATE: 24-JUN-1998  
CLASSIFICATION: 800

ATTORNEY/AGENT INFORMATION:  
NAME: Michaels, Christopher A  
REGISTRATION NUMBER: 34,390  
REFERENCE/DOCKET NUMBER: BTI-39  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (607) 256-3628  
TELEFAX: (607) 256-3628  
INFORMATION FOR SEO ID NO: 4:

SEQUENCE CHARACTERISTICS:  
LENGTH: 805 amino acids  
type: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
HYPOTHEICAL: NO  
ANTI-SENSE: NO  
ORIGINAL SOURCE:  
ORGANISM: Trichoplusia ni  
TISSUE TYPE: peritrophic membrane  
US-09-103-429A-4

Query Match 99.8%; Score 4352; DB 15; Length 805;  
Best Local Similarity 99.8%; Pred. No. 5.6e-271;  
Matches 805; Conservative 0; Mismatches 0; Indels 2; Gaps 2;

1 MIKLLFLTAAGLVAAAREVSDAKNPALHPHDPXPAEQXLLPXYDCTKFFYCYG 60  
1 MIKLLFLTAAGLVAAAREVSDAKNPALHPHDPXPAEQXLLPXYDCTKFFYCYG 60  
61 LKFIAPRDCAGTEFEKSAQCVAAALAGCTLPGPAAETTOAPATTOAPTTTQA 120  
61 LKFIAPRDCAGTEFEKSAQCVAAALAGCTLPGPAAETTOAPATTOAPTTTQA 120  
61 LKFIAPRDCAGTEFEKSAQCVAAALAGCTLPGPAAETTOAPATTOAPTTTQA 120  
121 APTTTOAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 180  
121 APTTTOAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 180  
121 APTTTOAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 180  
181 PTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 240  
181 PTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 240  
181 PTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQAAPTTTQA 240  
241 PICELLPNGCPADFDHLLPHDKYCNLFYQCSNGYFEORCPGLYFNRYVQCDSPAN 300  
241 PICELLPNGCPADFDHLLPHDKYCNLFYQCSNGYFEORCPGLYFNRYVQCDSPAN 300  
241 PICELLPNGCPADFDHLLPHDKYCNLFYQCSNGYFEORCPGLYFNRYVQCDSPAN 300  
301 VECDEGISAPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
301 VECDEGISAPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
301 VECDEGISAPAPVTEGNEDEDIDIGLLDNGCPANFEIDWLLPHGNRCDKYYQCVHGNLV 360  
361 ERRCGAGTHSFELQOCDHIELVAGCTLPGGESEVVDVEDACTGMYCTEPIEMEPPLNG 420  
361 ERRCGAGTHSFELQOCDHIELVAGCTLPGGESEVVDVEDACTGMYCTEPIEMEPPLNG 420  
361 ERRCGAGTHSFELQOCDHIELVAGCTLPGGESEVVDVEDACTGMYCTEPIEMEPPLNG 420









Db 382 ---PKKPAPTPKKEPAPTTTKKPAAPAPKKEPAPTPKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTK 438  
QY 536 -PAPTAAPTAAPTAAPESPTTVTP-PTAAPTAPTAAPTAAPESPTTVTP---SAPTAAPT 589  
Db 439 EKAPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTEPAPTE 496  
QY 590 AAPTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA 645  
Db 497 PAPTTPKKEPAPTPKKEPAPTTKKEP---APTTPKAPAPKELAPTTKKEPST-TSDKPAP 552  
QY 646 ---NTVTAPPTAAPT--AAPANTVTVPTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 694  
Db 553 TTKEGAPTTTPKKEPAPTPKKEPAPTTKGTAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 720  
QY 695 TTTS---APATPEDDDIDP--PLPNDPIMP 720  
Db 613 STTSKAPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTTKKEPAPTT 643

RESULT 7  
US-07-757-022B-2  
; Sequence 2, Application US/07757022B  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION NUMBER:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Gesner, Luann  
REGISTRATION NUMBER: 31,832  
REFERENCE/DOCKET NUMBER: GI 5190  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1404 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-2

Query Match 12.3%; Score 542; DB 2; Length 1404;  
Best Local Similarity 30.4%; Pred. No. 5,3e-27;  
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

QY 93 PGPPAE-TTQAPANT---QAPTTQAPTTT---QAPTTTQAPTTTQ-APTTO-AP 142  
Db 331 PTPAEPTTKGAPALTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKE 390  
QY 143 TTQQ---APTTOQAPTTT---QAPTTTQ-APTTOQAPTTTQ-APTTOQ-APTTOQ 192  
Db 391 TTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKE 450  
QY 193 PTTTQATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAAT 252  
Db 451 P---APTTPKEPTP---TPKKEPAPTTKKEPAPTTKKEPAP---APKPPA----- 492  
QY 253 DFDHILIPHDKCNLFYQCSNGYTFEQRCPEGLYFNPYQRODSPANVCDEISAP 312  
Db 492 -----PTTPKKEPAPTTK-----EPAPT 509  
QY 313 VTGNEDEDIDGLDNGCPANFEIDWLPNGNCKYQCVHGNLYERRCAGTHFSF 372  
Db 510 TTKE-----PS----- 516  
QY 373 ELQCCDHIELVGTLPGESEEVVDDEACTGWCPTPIEMBLP---NGCPADSIDH 429  
Db 516 -----PTTP---KKEPAPTTKSAPT----- 533  
QY 430 LPHESDCGYLQCVHGQIARPCGNLHSPANQSESVTACGVFEDSDNQCSTA 489  
Db 533 -----TTKKEPAPTTKSAPTTPKKEPST-----TKEP 560  
QY 490 APTA---APTAAPTAPTAAPTASTVVPATPATAPVPTTAAPTAPTAAPTAA 545  
Db 561 APTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKE 617  
QY 546 PTTAAPESPTT---VTVPPTAAPT---APTAVPEIPTVT---SAPTAAPTAPT 533  
Db 618 PTPPKLTPTTPKELAPTTTPKKEPAPTTPEELAPTT---PEBPTPTPEEPAPTTTPKKAAPN 675  
QY 594 AAPTAPTAAPTAAPTT---VTSPTAAPT-----AAPANT----- 629  
Db 676 TPKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAP 735  
QY 629 TTVPTAAPTAPAP---NTVTAPPTAAPT--AAPANTVTVPTAAPTAPPTV 683  
Db 736 TTTPKEPTST-TSDKPAPTTKGTAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 794  
QY 684 AH-----APNTAAPTPTTS---APATPEDDDIDP--PLPNDPIMP 720  
Db 795 KKPAPKELAPTTTKGPTSTSDKPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTTTPKKEPAPTT 842

RESULT 8  
US-07-757-022B-14  
; Sequence 14, Application US/07757022B  
; GENERAL INFORMATION:  
; APPLICANT: Gesner, Thomas G.  
; APPLICANT: Clark, Stephen C.  
; APPLICANT: Turner, Katherine  
; APPLICANT: Hewick, Rodney M.  
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
; NUMBER OF SEQUENCES: 143  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Genetics Institute, Inc.  
; STREET: 87 Cambridgepark Drive  
; CITY: Cambridge  
; STATE: Massachusetts  
; COUNTRY: U.S.A.  
; ZIP: 02140

COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



```
Db      PTPKAEHTTKGPALPTPKEPRTTPEKBASATPKEBPMTITKSAPTTTPKEBAPTTSAP    3478  
Oy     143 TTTO--AFTTQAPTTTT---QAFTTTQ-APTTTTAOPTTQ--APTTO--APTTOA    192  
Dd      348 TTPKEBAPTTPKREPATTTPKPBAPTTPTEKPAPTITSAPTTPKKABAPTTPKPAE    407  
Oy     193 PTITIQAATPATPATPATPATPATPATPATPGVAPTSPAVMPPICELLPNCGPA    252  
Dd      408 F-----ATTTKEPTR-TTPKEBAPTTPKEBAPTTPKEBAT-APRKA-----    449  
Oy     253 DFDHLLPHDKYCNLEYQCNSGTFTEORCEGLTENPVYORDCSPANVEDGEISAPP    312  
Dd      449 -----PTTPKEBAPTTPK-----EBAPT    466  
Oy     313 VTEGNEDDEDIDIGDLONGCPANFEIWLHPHNRCODKYVCVAGNLVERNCAGTHESE    372  
Dd      467 TTKE-----PS-----    473  
Oy     373 ELQQCDHIELVGCTLPGESEBEVDVEDACTGWYCPTETIEWEPLP--NGCPADFSIH    429  
Dd      473 -----PTTP--KEBAPTTPKSAF-----    490  
Oy     430 LLPHSDCGOYLQCVHGQTARPCBGNIHESPAQSCESPYTACQVFECDSNOGSTA    489  
Dd      490 -----TTKEBAPTTPKSAFTTPKESPST-----TRP    517  
Oy     490 APTA----APTAAPTAAPATAPTASTVVVPATPPAPAFAVPBTUALIPTAPTAFTA    545  
Dd      518 APPTTKEBAPTTPKRPAATTTPKREBAFTTPKRPA-PITTKRAP--TABKEBAPTTPETA    574  
Oy     546 PTTAABESPTT-----VTVPPTAAPT>--APTNAVPEIPITYT---SAPTAAPTAPT    593  
Dd      575 PTPPKKLPTTPPEKIAPTTPKEBAPTTPBELAPPT--PEESTPTTPKEBAPTTPKAAPN    632  
Oy     594 AAPTAAPTNAVPEIPT-----VISPPTAAPT-----AAPANT-----    629  
Dd      633 TPKEBAPTTPKEBAPTTPKEBAPTTPKETAPTTPKGTAFTTLKEBAPTTPPKRPAPLAP    692  
Oy     629 TVTVPTAPTAPTAAAP-----NTVTAPTAPTPT-AAAPENTTVTVPPTAAPTAPPTV    683  
Dd      693 TTTTKEBST-TSDRKAPTTPGTAPTTPKEBAPTTPKEBAPTTPKGTAFTTLKEBAPTTP    751  
Oy     664 AH-----APMPTAAPVTTS---APATPDDDIIDP--PLPNPIP    720  
Dd      752 KKPAPKELAPTTTGPISTSDDRKAPTTPKETAPTTPKEBAPTTPKRPR    799
```

RESULT 10  
US-07-757-022B-42  
Sequence 42, Application US/07757022B  
GENERAL INFORMATION:  
  APPLICANT: Gesner, Thomas G.  
  APPLICANT: Clark, Stephen C.  
  APPLICANT: Turner, Katherine  
  APPLICANT: Hewick, Rodney M.  
  TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
  NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
  ADDRESSEE: Genetics Institute, Inc.  
            Street: 87 Cambridgepark Drive  
            City: Cambridge  
            State: Massachusetts  
            Country: U.S.A.

ZIP: 02140  
COMPUTER READABLE FORM:  
      MEDIM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910

```

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/643,502
FILING DATE: 18-JAN-1991
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/546,114
FILING DATE: 29-JUN-1990
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/457,196
FILING DATE: 29-DEC-1989
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/390,901
FILING DATE: 08-AUG-1989
ATTORNEY/AGENT INFORMATION:
NAME: Csert, Luann
REGISTRATION NUMBER: 31,822
REFERENCE/DOCKET NUMBER: GI 5190
TELECOMMUNICATION INFORMATION:
TELEPHONE: (617)876-1170
TELEFAX: (617)876-5851
INFORMATION FOR SEQ ID NO: 42:
SEQUENCE CHARACTERISTICS:
LENGTH: 1311 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-757-022B-42

```

Query Match	13.3%	Score 542	DB 2	Length 1311
Best Local Similarity	30.4%	Pred. No. 5e-27		
Matches	215	Conservative	23	Mismatches 194; Indels 276; Gaps 37
QY	93	PGPPAE-ITQAPATY----	QAPITTOQAPITTY--QAPITTOQAPITTYO-APITTYO-AP	142
Db	238	PTPKAETTTKCPALATPKREPPTTKEKESATPKBEPPTTTSKAPATTPKEAPATTTKSA		297
QY	143	TTTTQ--APITTOQAPITTY--QAPITTYO-APITTOQAPITTYO--APITTYO--APITTOA		192
Db	298	TTPKAPATTTKEKAPATTPKKEPAPITTTKEKAPATTTTTSKAPATTTKEKAPATTPKAPATTPKE		357
QY	193	PTTTOQAATTPPAATTPPAATTPPAATTPPAATTPPAATTPGVAAPTSAVWPAPICELLFNGCPA		252
Db	358	P-----APITTKREPPT--TTKEKAPATTTKEKAPATTPKKEPAP--APKKA		399
QY	253	DFDHLILPHDKYCLFYQCSNGYFEEDRCBGLXFNYPVOCDSNPANVECCGEGISPAAP		312
Db	399	-----APITTKREPPT--TTKEKAPATTTKEKAPATTPKKEPAP--APKKA		416
QY	313	VTEGNEDEDIDIGLILDDNGCPANFEIDMLPHGNRODKYYQCVHGNLVERRCAGATHSEF		372
Db	417	TTKE-----PS-----		423
QY	373	ELQOCDHLEIVGCLTSGESESEVDVDEDACITGWGPTPEIEMWELP--NGCAPDESIDH		429
Db	423	-----PTTP--KEPATTTTKSAT-----		440
QY	430	LIPHESDCGYLQCVHGGTIANPCBGNLHFSFATGSCSPVTAAGCVFECDSNDQCTSTA		489
Db	440	-----TTKEPATTTTKSAPTTKEPSPIT-----TKER		467
QY	490	APTA-----APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA		545
Db	468	APITTKEPAPATTPKAPATTPKKEPAPITTKKAP--TAPKEPATTPKETA		524
QY	546	PTTAAPESPIT-----VVPPPTAPSTA-----APTAAPVEIPITVT-----SAPTAAAPTAAP		593
Db	525	PTPKKPLATTPPEKAPATTPKKEPAPITTPPELAPTT--PEEPPTPEDEAPATTPKKAAPN		582
QY	554	AAPTAAPTAAPEITTT-----VTSPPPTAAPT-----AAPAPNT-----		629
Db	583	TPKEAPATTPKKEPAPATTPKKEPAPATTPKETAATTPKGTATTTLKEKAPATTPKPKAPAKELAP		642



APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989  
PRIOR APPLICATION DATA: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Cseer, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEO ID NO: 46:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1320 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-46

Query Match 12.3%; Score 542; DB 2; Length 1320;  
Best Local Similarity 30.4%; Pred. No. 5e-27;  
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

93 GGPAAE-TTQAPATTT---QAPTTTQAPTTT---QAPTTTQAPTTTQ-APTTO-AP 142  
143 TTTO-APTTOAPT---QAPTTTQ-APTTOAPT---APTTO-APTTOA 192  
307 TTPKEAPPTTKKEAPPTTKKEAPPTTKKEAPPTTKSAPTTPKEAPPTTKKEAPPTTKKE 366  
193 PTTTQATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPA 252  
367 P-----APTTPKEPTP---TTPKEAPPTTKKEAPPTTKKEAPPT---APKKA----- 408  
253 DFDHLILPHDKYCNLYQCNGTFFGRCPEGILFNPYVGRCSNPANVEDGELISAPP 312  
408 -----PTTPKEAPPTTK-----EPAPT 425  
313 VTEGNEDEDIDIGDLNDCPANEIDWLPHGNRCDDKYQCVGNLYERRGAGTHESF 372  
426 TTKR-----PS----- 432  
373 ELQOCHIELVGTLPGESEEDVDDEACTGWCPTPEPIEMWLP---NGCPADFSDH 429  
432 -----PTTP---KEBAPTTPKSAPT----- 449  
430 LHPHSDCGYQVOCVHOTIARPCGNLHFSRATOSCEPYTACQVFECSNDGCTSTA 489  
449 -----TTKEPAPPTTKSAPTTPKEPSPT-----TKEP 476  
490 APTA-----APTAAPTAPTAAPTAPSTVVPATPPATAAPVPTTAAPTAPTAAPTAA 545  
477 APTTPKEAPPTTKKEAPPTTKKEAPPTTKKEAPPTTKKEAPPTTKKEAPPTTKKEAPPTTK 533  
546 PTTAAPSPT---VTVPPTAAPT---APTAVPEIDIVT---SAPTAAPTAPT 593  
534 PTPPKKLTPPEKLAFTPEKAPPTPEELAPTT---PEEPTPTTPEEPAATTPKAAPN 591  
594 AAPTAAPTAVPEIPT---VTSPTAAPT-----AAPAPT----- 629  
592 TPKEPAPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAPPTTPKEAP 651  
629 TVTPAPTAPTAAPT---NTVTAPTAPTAAPT---AAPANTTVTPAPTAPTAAPTAP 683  
652 TTTKEPPTST-TSDKPAPTTPKGTAPPTTPKEAPPTTPKEAPPTTPKGTAPPTTPKEAPPTTP 710  
684 AH-----APNTTAAVTTTS---APATTPDDDDIDP---PLPNDPIND 720

Db 711 KKPAPKEAPPTTKGPTSTSDKPAPTTPKEAPPTTPKEAPPTTPKEAPPTTPK 758

## RESULT 13

US-07-757-022B-48  
Sequence 48; Application US/07/757022B

## GENERAL INFORMATION:

APPLICANT: Gesner, Thomas G.

APPLICANT: Clark, Stephen C.

APPLICANT: Turner, Katherine

APPLICANT: Hewick, Rodney M.

TITLE OF INVENTION: Megakaryocyte Stimulating Factors

NUMBER OF SEQUENCES: 143

CORRESPONDENCE ADDRESS:

ADDRESSEE: Genetics Institute, Inc.

STREET: 87 Cambridgepark Drive

CITY: Cambridge

STATE: Massachusetts

COUNTRY: U.S.A.

ZIP: 02140

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/757,022B

FILING DATE: 19910910

CLASSIFICATION: 530

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/643,502

FILING DATE: 18-JAN-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/546,114

FILING DATE: 29-JUN-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/457,196

FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/390,901

FILING DATE: 08-AUG-1989

ATTORNEY/AGENT INFORMATION:

NAME: Cseer, Luann

REGISTRATION NUMBER: 31,822

REFERENCE/DOCKET NUMBER: GI 5190

TELEPHONE: (617)876-1170

TELEFAX: (617)876-5851

INFORMATION FOR SEO ID NO: 48:

SEQUENCE CHARACTERISTICS:

LENGTH: 1354 amino acids

TYPE: AMINO ACID

TOPOLOGY: linear

MOLECULE TYPE: protein

US-07-757-022B-48

Query Match 12.3%; Score 542; DB 2; Length 1354;  
Best Local Similarity 30.4%; Pred. No. 5.1e-27;  
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

93 GGPAAE-TTQAPATTT---QAPTTTQAPTTT---QAPTTTQAPTTTQ-APTTO-AP 142  
281 PTPKAEITTTGPAITTTKEPTTPKEPASTTPKEPTTPTKSAPTTPKEAPPTTKSAP 340  
143 TTTO-APTTOAPT---QAPTTTQ-APTTOAPT---APTTO-APTTOA 192  
341 TTPKEAPPTTKKEAPPTTKKEAPPTTKKEAPPTTKSAPTTPKEAPPTTKKEAPPTTKKE 400  
193 PTTTQATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPA 252  
401 P-----APTTPKEPTP---TTPKEAPPTTKKEAPPTTKKEAPPT---APKKA----- 442

QY 253 DEDHLLPHDRYCNLFYOCSSNGYTFEORCPGLYFNYPVORCDSPANECDGEISPAP 312  
Db 442 -----PTTKREPAPTTKK-----EPAPT 459  
QY 313 VTEGNEDEDIDGLDNDCCPANFEIDWLLPHGNRCDKYQCCHGNLYERRCAGTHSEF 372  
Db 460 TTKR-----PS-----TKER 510  
QY 373 ELQOCDHIEVGCTLRGSESEVDVEDACTGWCPTPEMEPLP---NGCPADFSIDH 429  
Db 466 -----PTTP---KEPAPTTKSAPT-----TKER 510  
QY 430 LPHESDCQYLQCVHQTIARPCGNLHFSPATOSCESPVTAGCOVFECDSNOCTSTA 489  
Db 483 -----TTKERAPTTTKSAPTTPKEPSPT-----TKER 510  
QY 490 APTA---APTAAPTAAPTAAPTAAPSTVPPATPATAPVPPTAIPTPAAPTAAPTA 545  
Db 511 APTTPKEPAPTTTPKPAAPTTTPKEPAPTTTPKEPAP---PTTKKPA---TAPKEPAPTTTPKE 567  
QY 546 PTTAPAESPTT-----VTVPPTAAPT---APTAVPEIPIVT---SAPTAPTAAPT 593  
Db 568 PTTTPKPLPTTPKEKLAAPTTPKEKAPTTPEBLAPT---PEEPTTPTEEPAPTTPKAAPN 625  
QY 594 AAPTAPTAAPTAAPTAAPTAAPTAAPTAAPTT-----AAPANT-----629  
Db 626 TTKERAPTTTPKPAAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 685  
QY 629 IVTVPPPTAAPTAPAP---NTVTAPPTAAPT---AAPANTVTVPPTAAPTAPPTV 683  
Db 686 TTKKEPTST-TSDKPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 744  
QY 684 AH-----APNTAAPTPTS---APATPEDDIDP---PLPNDP 720  
Db 745 KKPAPKELAPTTTKGPTSTSDKPAPTTPKEPAPTTTPKEPAPTTTPK 792

RESULT 14  
US-07-757-022B-50  
Sequence 50, Application US/07757022B  
GENERAL INFORMATION:  
APPLICANT: Gesner, Thomas G.  
APPLICANT: Clark, Stephen C.  
APPLICANT: Turner, Katherine  
APPLICANT: Hewick, Rodney M.  
TITLE OF INVENTION: Megakaryocyte Stimulating Factors  
NUMBER OF SEQUENCES: 143  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Genetics Institute, Inc.  
STREET: 87 Cambridgepark Drive  
CITY: Cambridge  
STATE: Massachusetts  
COUNTRY: U.S.A.  
ZIP: 02140  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/757,022B  
FILING DATE: 19910910  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/643,502  
FILING DATE: 18-JAN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/546,114  
FILING DATE: 29-JUN-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/457,196  
FILING DATE: 29-DEC-1989

PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/390,901  
FILING DATE: 08-AUG-1989  
ATTORNEY/AGENT INFORMATION:  
NAME: Csert, Luann  
REGISTRATION NUMBER: 31,822  
REFERENCE/DOCKET NUMBER: GI 5190  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (617)876-1170  
TELEFAX: (617)876-5851  
INFORMATION FOR SEQ ID NO: 50:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1314 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-757-022B-50

Query Match 12.3%; Score 542; DB 2: Length 1314;  
Best Local Similarity 30.4%; Pred. No. 5e-27;  
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

QY 93 PGPPAE-TTQAPAT---QAPTTQAPTTT---QAPTTQAPTTTQ-APTTO-AP 142  
Db 241 PTPKAPTTTKGALTTTPKEPTPTTPKEPASTTPKEPTPTTISAPTTPKEPAPTTTKSAP 300  
QY 143 TTTQ---APTTOAPT---QAPTTTQ-APTTOAPT---APTTO-APTTOA 192  
Db 301 TTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTKSAPTTPKEPAPTTTPKE 360  
QY 193 PTTTQATTPATPATPATPATPATPATPATPATPATPATPATPATPATPATPATPATPAT 252  
Db 361 P-----APTTPKEPT---TTPKEPAPTTTPKEPAPTTTPKEPAP---APKPA-----402  
QY 253 DEDHLLPHDRYCNLFYOCSSNGYTFEORCPGLYFNYPVORCDSPANECDGEISPAP 312  
Db 402 -----PTTKREPAPTTPK-----EPAPT 419  
QY 313 VTEGNEDEDIDGLDNDCCPANFEIDWLLPHGNRCDKYQCCHGNLYERRCAGTHSEF 372  
Db 420 TTKR-----PS-----TKER 470  
QY 373 ELQOCDHIEVGCTLRGSESEVDVEDACTGWCPTPEMEPLP---NGCPADFSIDH 429  
Db 426 -----PTTP---KEPAPTTKSAPT-----TKER 470  
QY 430 LPHESDCQYLQCVHQTIARPCGNLHFSPATOSCESPVTAGCOVFECDSNOCTSTA 489  
Db 443 -----TTKERAPTTTKSAPTTPKEPSPT-----TKER 470  
QY 490 APTA---APTAAPTAAPTAAPTAAPSTVPPATPATAPVPPTAIPTPAAPTAAPTA 545  
Db 471 APTTPKEPAPTTTPKPAAPTTTPKEPAPTTTPKEPAP---PTTKKPA---TAPKEPAPTTTPKE 527  
QY 546 PTTAPAESPTT-----VTVPPTAAPT---APTAVPEIPIVT---SAPTAPTAAPT 593  
Db 528 PTTTPKPLPTTPKEKLAAPTTPKEKAPTTPEBLAPT---PEEPTTPTEEPAPTTPKAAPN 585  
QY 594 AAPTAPTAAPTAAPTAAPTAAPTAAPTAAPTT-----AAPANT-----629  
Db 586 TTKERAPTTTPKPAAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 645  
QY 629 IVTVPPPTAAPTAPAP---NTVTAPPTAAPT---AAPANTVTVPPTAAPTAPPTV 683  
Db 646 TTKKEPTST-TSDKPAPTTPKGTAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 704  
QY 684 AH-----APNTAAPTPTS---APATPEDDIDP---PLPNDP 720  
Db 705 KKPAPKELAPTTTKGPTSTSDKPAPTTPKEPAPTTTPKEPAPTTTPK 752

RESULT 15



```

US-07-757-022B-52
; Sequence 52, Application US/07757022B
; GENERAL INFORMATION:
; APPLICANT: Geeser, Thomas G.
; APPLICANT: Clark, Stephen C.
; APPLICANT: Turner, Katherine
; APPLICANT: Hewick, Rodney M.
; TITLE OF INVENTION: Megakaryocyte Stimulating Factors
; NUMBER OF SEQUENCES: 143
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genetics Institute, Inc.
; STREET: 87 Cambridgepark Drive
; CITY: Cambridge
; STATE: Massachusetts
; COUNTRY: U.S.A.
; ZIP: 02140
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/757,022B
; FILING DATE: 19910910
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/643,502
; FILING DATE: 18-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/546,114
; FILING DATE: 29-JUN-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/457,196
; FILING DATE: 29-DEC-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/390,901
; FILING DATE: 08-AUG-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Cseer, Luann
; REGISTRATION NUMBER: 31,822
; REFERENCE/DOCKET NUMBER: GI 5190
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617)876-1170
; TELEFAX: (617)876-5851
; INFORMATION FOR SEQ ID NO: 52:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1363 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-07-757-022B-52

```

```

Query Match      12.38; Score 542; DB 2; Length 1363;
Best Local Similarity 30.44; Pred. No. 5.2e-27;
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

```

```

QY 93 GPPAE--TTOAPATT---QAPTTTQAPTTT---QAPTTTQAPTTTQ--APTTQ--APTTQ--AP
DB 290 PTPRAETTTGPRALTTPKEPTTPKEPASTTPKEPTTTKSAPTTPKEPAPTTTTSAP 349
QY 143 TTTQ--APTTQAPTTT---QAPTTTQ--APTTTQAPTTTQ--APTTQ--APTTQ--APTTQ 192
DB 350 TTPKEPAPTTTKEPAPTTTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 409
QY 193 PTTTQATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTP 252
DB 410 P-----APTPKEPTP--TTPKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKEPAPTTTKE 451
QY 253 DFDHLLIPHDKYNLFYQCSNGYTFQRCPEGLYFNPYVQRCDSPANVECDGELISAPP 312
DB 451 -----PTTPKEPAPTTTPK-----EPAPT 468

```

```

QY 313 VTGNEDEDIDLDLNGCPANFEIDWLLPHGNRCDKXYQCVHGNLVERRCAGTHFSF 372
DB 469 TKKE-----PS----- 475
QY 373 ELQOCHIELVGCTLPGESEBEVDVEDACTGWYCPTEPIEMEP--NGCPADSIDH 429
DB 475 -----PTTP--KEPAPTTTTSAPT----- 492
QY 430 ILPHESDCQGLQCVHGQTIARPCGNLHFSPATQSCSESVTAQCVFECDSNQCST 489
DB 492 -----TKKEPAPTTTTSAPTTPKEPSPT-----TKEP 519
QY 490 APTA---APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA 545
DB 520 APTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKE 576
QY 546 PTTAAPESPT-----VTVPPTAAPT---APTTAAPT---APTTAAPT---SAPTAPTAAPT 593
DB 577 PTPPKLTPTPPEKLAFTTPPEKAPTTPEELAPT--PEEPTTPPEEPAPTTPKAAAPN 634
QY 594 AAPTAAPTAVPEIPT---VTSPTAAPT-----AAPANT----- 629
DB 635 TPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 694
QY 629 TVTPPTAAPTAAAP---NTVTAAPTAPT---AAPANTTVTPPTAAPTAPT 683
DB 695 TTPKEPTST-TSDKPAFTTPPKGAPTTTPKEPAPTTTPKEPAPTTTPKEPAPTTTPKEPAP 753
QY 684 AH-----APNTAAPTPTS---APATPEDDDIDP--PLPNDPTNP 720
DB 754 KKPAPKELAPTTTGSTSTSDKPAFTTPKETAPTTPKEPAPTTTPKXP 801

```

Search completed: July 18, 1999, 03:46:52  
Job time: 1536 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 00:45:05 ; Search time 27.5 seconds

(without alignments)  
1099.291 Million cell updates/sec

Title: US-09-294-663-4

Perfect score: 4395  
Sequence: 1 MIKTLFLTALGLVARPEV.....VFIPNKKDEDIRQALNFEL 807

Scoring table: BLOSUM62

Searched: 116738 seqs, 37460341 residues

Database :

1: PIR-58:\*  
2: PIR1:\*  
3: PIR2:\*  
4: PIR3:\*  
5: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	660.5	15.0	3020	2	A43932	mucin 2 precursor,
2	460	10.5	873	2	A47283	calpain - fruit
3	457	10.4	865	2	A47282	calcium-binding pr
4	446.5	10.2	797	1	VCBEX1	glycoprotein X pre
5	430.5	9.8	662	2	A45155	mucin FIM-C.1 - Af
6	428	9.7	1335	2	A35887	mucin precursor, p
7	428	9.7	1255	2	B35175	episialin B - huma
8	428	9.7	1295	2	A35886	polymorphic epithe
9	428	9.7	1264	2	A35175	episialin A - huma
10	421.5	9.6	1188	2	S45915	extensin-like prot
11	410	9.3	1367	2	S48478	glucan 1,4-alpha-g
12	384	8.7	294	2	A37232	mucin, tracheal (A
13	374	8.5	822	2	A38420	antifreeze glycopr
14	373	8.5	1118	2	A48292	mucin, tracheobron
15	367.5	8.4	474	2	S15921	protein rpx-VT3 -
16	365	8.3	353	2	S36438	Eppit protein - hyd
17	352	8.0	670	2	F36791	mucin, hypothetical
18	348.5	7.9	279	2	S53363	mucin 5AC (clone J
19	347.5	7.9	350	2	S22456	gastric mucin (clo
20	345.5	7.9	528	2	I47141	larynx mucin (clon
21	345	7.8	232	2	A60095	larynx mucin (clon
22	341	7.8	303	2	S28264	hydroxyproline-ric
23	340.5	7.7	328	2	J00985	hydroxyproline-ric
24	340.5	7.7	216	2	I51920	mucin, rhesus mac
25	337	7.7	901	2	A42227	sialidase - Actino
26	336.5	7.7	267	2	S08314	cell wall glycopro
27	331	7.5	400	2	A28172	spasmolytic precu
28	324.5	7.4	379	2	S50125	larynx mucin (clon
29	324	7.4	1630	2	A53577	ascites sialoglyco
30	321	7.3	907	1	O0BE21	membrane antigen 9
31	321	7.3	376	2	S71558	mucin 7 precursor,
32	320.5	7.3	377	2	A46018	chitinase (EC 3.2.
33	318.5	7.2	860	1	JC4566	salivary glue prot
34	317	7.2	307	1	GSFF3	probable membrane
35	315	7.2	1161	2	S57180	glycoprotein 350/2
36	313.5	7.1	886	2	S29605	hydroxyproline-ric
37	311	7.1	620	2	S06333	hypothetical prote
38	309.5	7.0	548	2	E70546	promastigote surfa
39	309.5	7.0	327	2	S20074	

40	309	7.0	369	2	S20500	hydroxyproline-ric
41	308.5	7.0	283	2	S13383	hydroxyproline-ric
42	308.5	7.0	451	2	S71754	cellular hepatitis
43	308	7.0	322	2	A53715	apomucin precursor
44	306.5	7.0	447	2	A39321	mucin - rat (fragm
45	300.5	6.8	3149	1	O0BE8	BPL1 protein - hu

## ALIGNMENTS

RESULT 1  
A43932  
mucin 2 precursor, intestinal - human (fragments)  
N:Alternate names: mucin SMUC-41  
C:Species: Homo sapiens (man)  
C:Date: 10-Mar-1993 #sequence-revision 12-Apr-1996 #text-change 20-Mar-1998  
C:Accession: A43963; A45106; B45106; A43932; B33332; A61257; P00328; P00329  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Siddiki, B.; Kim, Y.S.  
J. Biol. Chem. 269, 2440-2446, 1994  
A:Title: Molecular cloning of human intestinal mucin (MUC2) cDNA. Identification of t  
A:Reference number: A43963  
A:Accession: A43963  
A:Molecule type: mRNA  
A:Residues: 1-639 <GU1>  
R:Gum Jr., J.R.; Hicks, J.W.; Toribara, N.W.; Roche, E.M.; Lagace, R.E.; Kim, Y.S.  
J. Biol. Chem. 267, 21375-21383, 1992  
A:Title: The human MUC2 intestinal mucin has cysteine-rich subdomains located both up  
A:Reference number: A45106; MUID:93016075  
A:Accession: A45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 626-1895 <GU2>  
A:Cross-references: GB:M94131; NID:9186395; PID:9186396  
A:Note: sequence extracted from NCBI backbone (NCBIP:116706)  
A:Accession: B45106  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 2037-3020 <GU3>  
A:Cross-references: GB:M94132; NID:9186397; PID:9186398  
A:Experimental source: colon  
A:Note: sequence extracted from NCBI backbone (NCBIP:116698)  
R:Toribara, N.W.; Gum Jr., J.R.; Cuihane, P.J.; Lagace, R.E.; Hicks, J.W.; Petersen,  
J. Clin. Invest. 88, 1005-1013, 1991  
A:Title: MUC-2 human small intestinal mucin gene structure. Repeated arrays and polym  
A:Reference number: A43932; MUID:91358717  
A:Accession: A43932  
A:Molecule type: DNA  
A:Residues: 1343-1350, 'L', 1352-1411, 'S', 1413-1448, 'P', 1450-1503, 'T', 1505-1915 <TOR>  
A:Cross-references: GB:M74027; NID:9188863; PID:9188864  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:55749; NCBIP:55750)  
R:Gum, J.R.; Byrd, J.C.; Hicks, J.W.; Toribara, N.W.; Lamport, D.T.A.; Kim, Y.S.  
J. Biol. Chem. 264, 6480-6487, 1989  
A:Title: Molecular cloning of human intestinal mucin cDNAs. Sequence analysis and evi  
A:Reference number: A33532; MUID:89197556  
A:Accession: B33532  
A:Molecule type: mRNA  
A:Residues: 1916-2193 <GU4>  
A:Cross-references: GB:M22405; NID:9188873; PID:9188874  
A:Experimental source: intestine  
R:Iany, B.H.; Gallup, M.W.; Yan, P.S.; Gum, J.R.; Kim, Y.S.; Basbaum, C.B.  
J. Clin. Invest. 87, 77-82, 1991  
A:Title: Human bronchus and intestine express the same mucin gene.  
A:Reference number: A61257  
A:Accession: A61257  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 'T', 1925-1948, 'TTS', 1952-1954 <JAN>  
A:Experimental source: bronchus  
R:Xu, G.; Huan, L.; Khatri, I.; Sajjan, U.S.; McCooll, D.; Wang, D.; Jones, C.; Forstn  
Biochem. Biophys. Res. Commun. 183, 821-828, 1992

Query Match	15.0%	Score 660.5;	DB 2;	Length 3020;
Best Local Similarity	26.1%;	Pred	No.1.5e-25;	
Matches 210;	Conservative 41;	Mismatches 204;	Indels 349;	Gaps 27

```

Db 1751 TTSSLPTTTPLPSPITPTPTSPSTPTPTTPTTCTCPLCMNTGMLDSGKRPFHNRGGDTIELG 1810
QY 749 VVCSGELQFNPTTTCDFACNVC 772
      |||
Db 1811 DVCGPG-----WAANISC 1823
      :|:|

RESULT 2
A47283
calphotin - fruit fly (Drosophila melanogaster)
C:Species: Drosophila melanogaster
C:Date: 21-Sep-1993 #sequence_revision 25-Apr-1997 #text_change 10-Sep-1997
C:Accession: A47283
R:Ballinger, D.G.; Xue, N.; Harshman, K.D.
Proc. Natl. Acad. Sci. U.S.A. 90, 1536-1540, 1993
A:Title: A Drosophila photoreceptor cell-specific protein, calphotin, binds calcium a
A:Reference number: A47283; MUID:93165730
A:Contents: photoreceptor cells
A:Accession: A47283
A:Status: preliminary
A:Molecule type: nucleic acid
A:Residues: 1-873 <BAL>
A:Cross-references: CB:L05080; NID:g157071; PID:g157072
A:Note: Sequence extracted from NCBI backbone (NCBIN:124958, NCBIPI:124959)

```

	Query Match	10.5%	Score 460;	DB 2;	Length 873;	
	Best Local Similarity	24.1%	Pred. No. 3.8e-16;			
	Matches 198;	Conservative 45;	Mismatches 251;	Indels 326;	Gaps 35;	
OY	91 TLPGPAAETTOAPAM--TQAPTTOAFTTTTOAFTTTTOA-----	130				
Dd	5 TIPSPVSAVAAPATPSAAVAPGVVSAAVAAVAPAAVAPAAPTAVTPAAPPTLASV	64				
OY	130 ---TTTTAP-----TTQAPPTTQAFTTTOAFTTTTQAFTTTTQAFTTQAFTTQAFT	182				
Dd	65 QPATVTAPAPAPIAASVASVAPVAAAPVAAATPPAASPST--PPVVAQIIPVAVSAPV	122				
OY	183 TQOAPTTOAQFTTTIQ--AATTPAATTP---AATTPAATTPAATTPTATTPGVPASTAP	237				
Dd	123 ---APVAAATTPPAVIPVAAVPIATTPVAAASATPPAAVTIVSVIAITPPVPANTTVP	179				
OY	238 VMPRICELLPMGCPADDEDIHLLIPHDKYCNLFYCNSNGYTEORCEGLTENPVYQRCDS	297				
Dd	180 VAAPVA-AVPAAPV-----	203				
OY	298 PAAVECDGEISBPAPVTEGENDEDDIDICDLDNCGPANFEIDWLLPHGNKCDKYVCVHG	357				
Dd	204 PAAPVAVPETAPRPVAA--IPVATIPECVAP-----LIEVYS-----	240				
OY	358 NLVERCAGAHFSEELQQCHIELYCGTLGEGSEEDVDVEDACTGYCCPREITEMEPL	417				
Dd	240 -----VATKRLPAAEPVVAA-PRA-----TE-----T	261				
OY	418 PNGCADFSIDLHPHESDCQYLQCVGQFIARPCPNLHFSPATOCESPVTAGCOVF	477				
Dd	262 PVVAPAAS--PHVS-----VAPAVTAAYAPASATEPPVAA-----	298				
OY	478 ECDSDNOCTSTAAPTAAPTAA-----PTAAPTAASTVVP-----	518				
Dd	298 -----ATLTTPA-ETPDALAPVVAESQVAAANTVATPTTPA--EPBTIAPPVAAETPE	347				
OY	518 -----ATPPATAAPVPPTTALTTPAPTAAPTAAPTAADESTTYT----VPPTAA	564				
Dd	348 VASVAVAEETPPV----VPVAAESIAPAVVATTPVPTLATVTDVDTASAVELELPVIA	403				
OY	565 PTAAPTETVPELPIV-----TAPTAAPT-----	590				
Dd	404 PSPVP-SAAVETPYDLAPVLCPVAAEEVPVAVAEETPETPAPASAPVTIALDIPEVAP	462				
OY	590 --AAPTAAPTAAPTAAVEI---PTTVTSPTTAAPTTAA-----AP	626				
Dd	463 VTAASDAPAEKPSMAAIIVSTPTTAAVPEETAPPAVPPEPIDIVSLSEAAITETPAP	522				

QY 627 NTTVT-----VPTAA-----PTTA 641  
DB 523 PVEVTEVAVADVAPPEAEDLIEVEPPAPIDLEOTTSPAVEAESTSSIPETS 582  
QY 642 AAPANTTVAPPTA-APTTA---APANTTVYPTTAAPPTAAPTAAHAPNTTAAV--T 695  
DB 583 LPPNNAVASPEVAVAPITAPERIPERPSLATPTPEVAPVIOEAVDAVEVTEET 642  
QY 696 TTSAPATPE-----DDIDPLPNDPI---NPCVEECN 726  
DB 643 STSIPETIVEFPEAVAEKVLDAITAPYTTOEDVYANIN 682

## RESULT 3

Calcium-binding protein calphostin - fruit fly (Drosophila melanogaster)  
A:Accession: A47282  
C:Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 20-Mar-1998  
R:Martin, J.H.; Benzer, S.; Rudnicka, M.; Miller, C.A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 1531-1535, 1993  
A:Title: Calphostin: a Drosophila photoreceptor cell calcium-binding protein.  
A:Reference number: A47282; MUID:93165729  
A:Accession: A47282  
A:Status: preliminary  
A:Molecule type: nucleic acid  
A:Residues: 1-665 <MAR>  
A:Cross-references: GB:102111; NID:9157031; PID:9157032  
A:Experimental source: photoreceptor cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:124955, NCBI:124956)  
C:Genetics:  
A:Gene: FlyBase:Cpn  
A:References: FlyBase:FBgn0010218  
C:Keywords: calcium binding

Query Match 10.4%; Score 457; DB 2; Length 865;  
Best Local Similarity 24.2%; Pred. No. 5.3e-16;  
Matches 197; Conservative 46; Mismatches 250; Indels 320; Gaps 35;

QY 91 TLPGPAETTOAPAT---TOAPTTQAPTTTQAPT-----TTTQAPTTT 133  
DB 5 TLSPVASAVAPVTPSAVAPQVAVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 64  
QY 134 QAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 189  
DB 65 PAP 118  
QY 190 TQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 244  
DB 119 AAPPTPVVQIPVAP 178  
QY 245 LIPNCPADFDLHLIPHDKICNLFYQCSNGTYEORCEGLXENPYVORCDSPANVED 304  
DB 178 AVPAVAVP-----VAVPAVAVP-----VAVPAVAVP 202  
QY 305 GELSPAPVTEGNEDEDIDGLDNGCPANFEIDLPHGNCKDXYQCVRHNLVERRC 364  
DB 203 AETPAPPAE-----IPVATIPCEVAP-----LIPVAVS----- 232  
QY 365 GAGTHSEFELQOCHIELVGLTLPGESESEVDVEDACTGWCPTPIEMEPINCCPAD 424  
DB 232 -----VATKPLAAAPVAVAPAPAPAPAPAPAPAPAPAPAPAPAPAPAPAP 260  
QY 425 FSLDLHPHESDCGOYLQCVHGOCTARPCPGNLIHSPAFQSCSPVYACQVFECDSDNO 484  
DB 261 AS-----PHVS-----VAPAVETAVVAVPVSASTEPVPA----- 290  
QY 485 CSTAPTAAPTAA-----PTAAPTAA PSTVVP----- 518  
DB 290 ATLITTP-ETPALAPVAVASOVANNTVATPTPAP--EBETIAPVVAETPEVASAVA 346

QY 518 -ATPPATAVPPTTAPPTAPPTAPPTAPPTAPPTAPPTAPPTAPPTAPPTAPPT 571  
DB 347 ETTPPV-----VPPVAESIPAPVAVATTPVAPATLAATDPPVATASAVDELPPVAPSVP-S 401  
QY 572 AVPEIPITY-----TSAPTAPT-----AAPT 594  
DB 402 AAETIPVDLAPVLPVPAEVPVAVAEETPEPTAPASAPVITIALDIPVAPVIAPSD 461  
QY 595 APTAAPTAVPEI-----PTVTSPTTAPPTTAP-----APNTTVT-- 632  
DB 462 APAPASAPAPVSTPTTASVETTAPAPAVPTERIDVSVSEAIETPVAPPEVITTE 521  
QY 632 -----VPTAA-----PTTAAAPNTT 648  
DB 522 VAVADVAPPEAAADLIEVEPPAPIDLEOTTSPAVEAESTSSIPETSIPPNNA 581  
QY 649 VAPAPTA-APTTA---APAPNTTVYPTTAAPTAPPTAAHAPNTTAAV--TTTSAPT 702  
DB 582 VASPEVAVAPITAPERIPERPSLATPTPEVAPVIOEAVDAVEVTEETSTISIPET 641  
QY 703 TPE-----DDIDPLPNDPI---NPCVEECN 726  
DB 642 TYEFPEAVAEKVLDAITAPYTTOEDVYANIN 674

## RESULT 4

## VGBEX1

Glycoprotein X precursor - equine herpesvirus 1 (strain Ab4p)  
C:Species: equine herpesvirus 1  
A:Note: host Equus caballus (domestic horse)  
C:Date: 30-Sep-1992 #sequence\_revision 30-Sep-1992 #text\_change 05-Sep-1997  
A:Accession: H36802  
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.  
submitted to GenBank, March 1992  
A:Description: The DNA sequence of equine herpesvirus-1.  
A:Reference number: A36805  
A:Accession: H36802  
A:Molecule type: DNA  
A:Residues: 1-797 <TEL>  
A:Cross-references: GB:M6664; NID:9330791; PID:9330862  
R:Telford, E.A.R.; Watson, M.S.; McBride, K.; Davison, A.J.  
Virology 189, 304-316, 1992  
A:Title: The DNA sequence of equine herpesvirus-1.  
A:Reference number: A41831; MUID:92295566  
A:Contents: annotation; possible protein-coding frames  
A:Note: neither amino acid nor nucleotide sequence is given  
C:Genetics:  
A:Gene: 71  
C:Superfamily: equine herpesvirus glycoprotein X; equine herpesvirus 1 glycoprotein h  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-797/Product: glycoprotein X #status predicted <MAT>  
F:23-465/Region: serine/threonine-rich  
F:489-797/Domain: equine herpesvirus 1 glycoprotein homology <EHG>  
F:766-790/Domain: transmembrane #status predicted <TMN>  
F:590/Binding site: carbohydrate (asn) (covalent) #status predicted

Query Match 10.2%; Score 446.5; DB 1; Length 797;  
Best Local Similarity 25.3%; Pred. No. 1.6e-15;  
Matches 182; Conservative 41; Mismatches 186; Indels 311; Gaps 20;

QY 78 SAOTCVHALLACTLPGPAETTOAPPTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTT 135  
DB 77 STOSSSTATSS-SAPSTASTSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 135  
QY 136 PTT-----TOAPTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTT 172  
DB 136 VTTASTASTTATATATATSTPTTTTATTTTATTTTATTTTATTTTATTTATTT 195  
QY 173 QAPTTTQAPT---TQAPTTTQAPT---TQAPTTTQAPT---TQAPTTTQAPT---TQAPTT 223  
DB 196 TATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 255



Best Local Similarity 26.78; Pred. No. 2e-14;  
Matches 208; Conservative 45; Mismatches 294; Indels 232; Gaps 37;

OY	17	REVSDAENKPNALHE--PHRDPRAPOXHLPEXEDCTFFYCEGLKFIARDCAPGTE	74
Db	105	RHALOST--TPRAHVYTSAPDNKRPAGSTAPPA-----HGVSAPDTRRAPGS-	151
OY	75	EKFSATOCVHAALACCTLGPRAETTOAPATQAFTTTOAFTT-----TOAFTTTO	127
Db	151	---TAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRA	207
OY	128	APTTOAFTTTOAFTTTOAFTTTOAFTTTOAFTTTOAFTTTOAFTTTOAFTTTOA	186
Db	208	PGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRA	267
OY	167	PTTQAFT--TIQAFTT---PAATTAA---TTPAATTAAATTAAATTGVAFTSAP--	238
Db	268	PGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRA---PGSTAPRAHGVSADPT	324
OY	238	-----WPPRICELLPNGCPADFDIHLIPHDKYCNLFYQCSNGYTF--EQGCPGLTFN	289
Db	325	RPAGSTAP-----AHGVTSAADPTRAPAGSTAP	353
OY	290	PVQRCDSPPANVECDGEISAPBPVTEGNEDEDIDIGILLDNGCPANFEIDWLLPHGNCD	349
Db	354	PAHGVTSAADPTRAPAG--STAPP-----	375
OY	350	KYYQCVHGNLVRRGAGTHFSFELQOCDHIELVCTLPBGSESEVDYDEDACHTWCP-	409
Db	375	---AHG-----VTSADPTRAPAGSTAPRAHGVSADPTRAPAGSTAP	414
OY	409	---TEPIEMEBLPNGCPADFSIDLHPHESDCGOYLACVHGOTIA---RPGGNLHRSP	461
Db	415	AHGVTSAADPTRAP-----GSTAPRAHGVTAPDTRRAPGST-APP	454
OY	462	ATQSCSEPVTAQGVFECDDSNQCTSTAAPTAAFTAAPTAAFTAAPTAAFTV--PEAT	519
Db	455	AHGVTSAAP-----DTRAPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPT	504
OY	520	PPATAAPVP---TTAIPPAFTAAPTAAFTAAESFTYIVPPTAAFTAAPTAAFTAVEI	576
Db	505	RPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPT	564
OY	577	PIYTSAP---TAAFTAAPTAAFTAAPTAAFTAVEIPTT-----VTASP---TAAFTT	621
Db	565	RPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPT	624
OY	622	AAPABNTTVVPP---TAAPTTAAPABNTTVTAP---TAAPTTAAPABNTTVTVP	673
Db	625	-RPAGGS--TAPRAHGVSADPT-RPAGGS--TADPAHGVSADPT-RPAGGS--TAPRA	675
OY	674	AAPTAAPT-----VAAPNTTAPVTT-----TSAPNTPEDDIDDP	712
Db	676	HGVTSAPDTRAPAGSTAPRAHGVSADPTRAPAGSTAPRAHGVSADPTRAPAGSTAP	734

Query Match	9.78; Score 428; DB 2; Length 1255;
-------------	-------------------------------------

	Best Local Similarity	Pred. No. 1.9e-14;	Matches 208; Conservative	45; Mismatches 294;	Indels 232;	Gaps 37;
	26.78;					

```

0Y      17  REBDAKKNALHE--PHRDXPRAEOXLLBPXYEDCKFYCCGYGLKFIAPRCAGTE  74
Db      105  RALLST---TPRAHGVISADPNKRPAEGSTAPRA-----HGVTASDTRPAPGS-  151
0Y      75  FKEFSQTCVHAALAGCULPRGPAETTOAPTATQAPTTQAETTT-----TOAFTTQQ.127
Db      151  ---TAPRAHGVTSADPTRAPAGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPTRPA  207
0Y      128  APTTTQAPTTQAPTTQAPTTQAPTTQAPTTQAPTTQAPTTQAPTTQAPTTQAPTTQA  186
Db      208  POSTAPRAHGVTSADPTRAPAGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPTRPA  267
0Y      187  PTTQAAPT-TTQAATT---PAATTBRA---TTAAITTPAATTAAITTPCAATPSAP--  238
Db      268  PGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPTRPA---PGSTARPAHGVTSADPT  324
0Y      238  -----WMPRIICELLPNGCPADFIDILLIPHDKYCNLFYCQSNQYTF--EQRCEGLYFN  289
Db      325  RPABESTAP-----AHGVTSADPTRRAPGSTAP  353
0Y      290  PYVQRCDSPANVECDGELISAPRPYTEGNEDDEDIDIGDLLNCGPANFEIDLPHGNKCD.  349
Db      354  PAHGVTSADPTRRAPG--STAPR-----  375
0Y      350  KYQCQVHNLYERRCGASTHSFELQCCDHILYGCILPGSESEVEDVDACGTWCYCP-  409
Db      375  ----AHG-----VTSADPTRAPAGSTARPAHGVTSADPTRRAPGSTAP  414
0Y      409  ----TERIEWERLPRNGCPADFSIDHLPLRHESDCQOYLQCVGOTIA---RPCPNLHSP  461
Db      415  AHGVTSADPTRAPR-----GSTAPRAHGVTSADPTRRAPAGST--APP.454
0Y      462  ATQSCSEPVTAGCOVFECDSNOCTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTVV--PPAT  519
Db      455  AHGVTSAP-----DTRAPAGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPT  504
0Y      520  PPATAAPRP---TTAITPRAPTAAPTAAPTAAPDESTTVYTVPRTAAPTAAPTAVPEI  576
Db      505  RPAPGSTAPRAHGVTSADPTRAPAGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPT  564
0Y      577  -PIVTSAP-----TAAPTAAPTAAPTAAPTAVPEIPT-----VTSPP---TAAPTPT  621
Db      565  RPAPGSTAPRAHGVTSADPTRAPAGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPT.  624
0Y      622  AAPAPANTVTVPP---TAAPTTAAAPANTVTVAPP---TAAPTTAAAPANTVTVAPT  673
Db      625  --RPAGS--TAPRAHGVTSADPT-RPAGS--TAPRAHGVTSADPT-RPAGS--TAPRA  675
0Y      674  AAPTAAPT-----VAHAPNTTAAPVTT-----TSAPATTPPEDDDIDDP  712
Db      676  HGVTASADPTRAPAGSTARPAHGVTSADPTRAPAGSTARPAHGVTSADPTRAPAGSTARP  734

```



A:Reference number: S49915  
A:Accession: S49915  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-1188 <RUB>  
A:Cross-references: EMBL:Z34465; NID:g600117; PID:g600118

Query Match 9.6%; Score 421.5; DB 2; Length 1188;  
Best Local Similarity 22.7%; Pred. No. 3.8e-14;  
Matches 175; Conservative 89; Mismatches 323; Indels 183; Gaps 29;

QY 2 IKTLLFLTALGLVAARPEVSDAEKNPALHEPHDX-----PPAEQXXLLLPXEDYDCTKFFY 56  
DB 520 VKTSPAPIGSPSPVPVSVSPVVKSPPPAPVGVSPPPPEKSPPPP-----570  
QY 57 CEYGLKFIAPRDCAPCTEFKFAQTCVHAALAGCTLPGPAAETQAPATTQAP-----110  
DB 570 -----APVASPPPVKSPPTLVASPPPVKSPPPAPVASPPPVKSPPTTPVA 621  
QY 110 -TTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 168  
DB 622 SPAPPAPVASSPPPMKSPPTPVSPPPPEKSPPPPPAKSTPPPEEYPTPTTSVKSSP 681  
QY 169 TTTQAPTTTQAP-----TTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 215  
DB 682 PPEKSLPPPTLTPSPPOEKPTPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSPSP 741  
QY 215 -----AATPAATTAAATPGV-----PAPTSAPVWPPICELLPNCPCADFDHLLIPH 262  
DB 742 SSPPTPVSSPPALAPVSPSPVVKSPSPPPAPLSSP--PPAPQV--KSSPPPVQVSSPPPA 797  
QY 263 DRYCNLFYQCSNGYTPERC-PEGLYFNFYVORCDSPANVECDGEISPAAPPVTEGNEDED 321  
DB 798 PKSSPLAPVSSPPQVEKTSPPAPLSSPLAPKSSPPHV---VVSSPPPVKSSPP--852  
QY 322 IDIGDLLNGCPANFIDMLPHGNRCNDKYYQCVHGNLVERCGAGTHFSFELQOCDHIE 381  
DB 852 -----PAPVSSPLTP-----KPASPPAHVS-----SPPE 876  
QY 382 LVGCTLPGESEVDVEDACTGWYCTPTPIEWELPNCPCADFDHLLPHESDGOVL 441  
DB 877 VKPSTPPAPTIVISP-----PSEKSSPP-----PTPVSLPPPIVKSS-----916  
QY 442 QCVHGOTIARPCGNLHFSPAT-QSCSESVTAGCQVFECDSDNQCSTTAAPTAAPTAAPT 500  
DB 916 -----PPPAMVSSPPMTPKSSPPVWV-----SSPPPTVKSSPPPPAPVSSPP 957  
QY 501 AAPTAAAPTAASTVVP--ATPPATAAPVPTTAPTAPTAAAPTAAAPTAAAPTAAAPT 557  
DB 958 ATPKSSPPPPAPVNLPPPEVKSSPPPTPVSSPP-----PAPKSSPPPPAPVSSPP--PPEV 1009  
QY 558 TVPPTAAAPTAAV--PEPIIVTSAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 613  
DB 1010 KSGPPPPAPVSSPPPPVVKSPPPAPVSSPPPPVVKSP--PPAPVSSP-----PPPVKS 1059  
QY 614 PPTAAPTAAAPNTVTVPPPTAAPTAAAPNTTVPPTAAPTAAAPTAAAPNTTVPPT 673  
DB 1060 PPPAPISPPPP--VKSPPPAPVSSPPPP--VKSPPPAPVSSPPPP--VKSPPPAPVSSPPPP--1110  
QY 674 AAPTAAPTVAHAPNTAAPTVTTTS-APATT-----EDDDIDPPLPNDP 717  
DB 1111 PAPVSSPPPPAPVKKPSLPPPPAPVSSPPPVVTPAPPKKEQSLPPPAESQP 1160

RESULT 11  
S48478  
N:Alternate names: extracellular glucosylase; mucin-like protein MUC1; protein Y1R019C  
C:Species: Saccharomyces cerevisiae  
C:Date: 02-Dec-1994 #sequence\_revision 02-Dec-1994 #text\_change 05-Dec-1997  
C:Accession: S48478; A26877; S27281; JC6123  
R:Rowley, K.

submitted to the EMBL Data Library, October 1994  
A:Reference number: S48478  
A:Accession: S48478  
A:Molecule type: DNA  
A:Residues: 1-1367 <ROW>  
A:Cross-references: GB:Z47047; EMBL:Z38061; MIPS:Y1R019C; NID:g603997; PID:g763364  
A:Yamashita, I.; Nakamura, M.; Fukui, S.  
J. Bacteriol. 169, 2142-2149, 1987  
A:Title: Gene fusion is a possible mechanism underlying the evolution of STA1.  
A:Reference number: A91831; MUID:87194600  
A:Accession: A26877  
A:Molecule type: DNA  
A:Residues: 1-242 <YAM>  
A:Cross-references: EMBL:M16164; NID:gl72522; PID:gl72525  
A:Accession: B26877  
A:Molecule type: DNA  
A:Residues: 762-1331 <YAZ>  
A:Cross-references: EMBL:M16165; NID:gl72523; PID:gl72526  
R:Pardo, J.M.; Ianez, E.; Zalacain, M.; Claros, M.G.; Jimenez, A.  
FEBS Lett. 239, 179-184, 1988  
A:Title: Similar short elements in the 5' regions of the STA2 and SGA genes from Sacc  
A:Reference number: S27281  
A:Accession: S27281  
A:Molecule type: DNA  
A:Residues: 1-31 <PAR>  
A:Cross-references: EMBL:X13857; NID:g4551; PID:g4552  
R:Lambrechts, M.G.; Bauer, F.F.; Marmur, J.; Pretorius, I.S.  
Proc. Natl. Acad. Sci. U.S.A. 93, 8419-8424, 1996  
A:Title: Muc1, a mucin-like protein that is regulated by Mss10, is critical for pseud  
A:Reference number: JC6123  
A:Accession: JC6123  
A:Status: nucleic acid sequence not shown  
A:Molecule type: DNA  
A:Residues: 1-1367 <LAM>  
A:Cross-references: GB:U30626; NID:gl304386; PID:gl304387  
C:Genetics:  
A:Gene: SGD:MUC1; STA2; MAL5; DEX2  
A:Cross-references: MIPS:Y1R019C; SGD:S0001458  
A:Map position: 9R  
C:Superfamily: yeast glucan 1,4-alpha-glucosidase homolog; glucan 1,4-alpha-glucosida  
C:Keywords: glycosidase; hydrolase; polysaccharide degradation; transmembrane protein  
F:5-21/domain: transmembrane #status predicted <TM1>  
F:1350-1366/domain: transmembrane #status predicted <TM2>

Query Match 9.3%; Score 410; DB 2; Length 1367;  
Best Local Similarity 19.9%; Pred. No. 1.6e-13;  
Matches 152; Conservative 124; Mismatches 238; Indels 248; Gaps 28;  
QY 32 PHPDXPAPAEQXXLLPXEYDCTKFFYCEYGLKFIAPRDCAPCTEFKFAQTCVHAALAGCT 91  
DB 288 PHHDTP-----CTK-----KKTTTCTKTKTTTPVTPSSSTTESSAP 328  
QY 92 LPGPPAETQ---APATQAPTTTQAPTTTQAPTT-----TQAPTTTQAPTTTQAP 142  
DB 329 VTPSSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPVTSSTTESSAPVTSST 388  
QY 143 TTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTT-----TQAPTTTQAPTTTQAPTT- 196  
DB 389 ESSAPVTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTS 448  
QY 196 -TQAPTTAAPTAAATTAAATTP--AATTAAATTPGPAP-----TSAPVWPPICEL 245  
DB 449 STTESSAPVTPSSSTTESSAPVTSSTTESSAP--VTPSSSTTESSAPV-----501  
QY 246 LPNGCPADFDHLLPHDKYCNLFYQCSNGYTFEQRCEGLYFNFYVORCDSPANVECDG 305  
DB 501 -----TSSTTESSAP-----VTPSSSTTES 522  
QY 306 EISAP-PVTEGNEDEDIDIGDLLNGCPANFIDMLPHGNRCNDKYYQCVHGNLVERRC 364  
DB 523 SSAPAPTSSSTTESSAPV-----543



QY 365 GAGTHSFELQOCDHIELVCGTLPGESESEVDVDEDACTGWCPTPIEWEPLNCGPAD 424  
Db 543 -----TSSSTESSAPV-----PTPSSSTTE 563  
QY 425 FSIDHLLPHESDCGQYLOCVHGQTIARPCGNLHFSFATOSCEPVYTAGQVFECDSDNQ 484  
Db 564 SS-----STPV-----SSTESSAPVPT-----PSSST 588  
QY 485 CTSTAAPTAA-----TAAPTAAPTAAAPSTVVPAPPTAPPAAPV-----PT 530  
Db 589 TESSAPVPTSSSTTESSAPAPTSSSTESSAPVPTSSSTESSAPVPTSSSTES 648  
QY 531 TAIPTAAPTAAAPTAAPESTTVVPTTAAPTAAAPTAAVPEIPIT-VTSAPTAAPT 589  
Db 649 SSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTESSAPV 707  
QY 590 AAPTAAAPTAAVPEIPITVTSPTAAAPTAAAPANTVITV-----VPTTAAPTAAAP 645  
Db 708 TPTSSSTTESSAPVPTSSSTTESSAP-----VPTPTSSSTTESSAPVPTSSSTESSAP 763  
QY 646 NTWTAPPTAAAPTAAAPNTVITV-----VP-PTAAPTAAAPTVAHAPNTTAAAPVITTSAP 700  
Db 764 VPTPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTTESSAPVPTPTSSSNITSSAP 823  
QY 701 ATP-----EDDDIDPLND-----PINPCVECNVLP 729  
Db 824 SSTPTSSSTESSVPVPTPTSSSTTESSAPVPTSSSTTESSAPVPTSSSTESSAPV 865  
RESULT 12  
A37232  
C:Species: human (fragment)  
C:Date: 14-Feb-1992 #sequence\_revision 14-Feb-1992 #text\_change 30-Sep-1993  
C:Accession: A37232  
R:Gerard, C.; Eddy Jr., R.L.; Shows, T.B.  
J. Clin. Invest. 86, 1921-1927, 1990  
A:Title: The core polypeptide of cystic fibrosis tracheal mucin contains a tandem repeat  
A:Reference number: A37232; MUID:91072667  
A:Accession: A37232  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-294 <GER>  
A:Cross-references: GB:M57417

Query Match 8.7%; Score 384; DB 2; Length 294;  
Best Local Similarity 24.1%; Pred. No. 8.4e-13;  
Matches 141; Conservative 8; Mismatches 132; Indels 304; Gaps 15;

QY 100 TQAPATTQAPTQAPTITQ-----APT-TTQAP-----TTTQAPTQAPTQAPT 150  
Db 3 TPTPTGTQPTPTPTTTTWTPTPTITSTQPTPTPTPTTTTWTPTPT--PSTQRTTP 60  
QY 151 TQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPT 209  
Db 61 TSITTTTWTPTPT-----PITTTTWTPTPTGTQPTPTPTPTPTPTPTPTPTGTL 117  
QY 210 ATPAATTAAPTTAAPTTGPVAPTSAPVWPPICELLNPGCPADFIDHLLPHDKYCNLF 269  
Db 118 TPTPTPTT-TTPTPTPTGTQPTPTPT----- 146  
QY 270 YQCSNGYFEQRCPEGLYFNYPVQRCDSANVEDGEISAPPVTEGNEDEDIDGLLD 329  
Db 146 ----- 146  
QY 330 NGCPANFEIDWLLPHGNRCDKYYQVHGNLVERRCGAGTHFSFELQOCDHIELVGCTLP 389  
Db 146 ----- 146  
QY 390 GESEVDVDEDACTGWCPTPIEWEPLNCGPADFSIDHLLPHESDCGQYLOCVHGQTI 449  
Db 146 ----- 146

QY 450 ARPCPNLHFSFATOSCEPVYTAGQVFECDSDNOCTSTAAPTAAAPTAAAPTAA 509  
Db 146 -----TTTWTPTPTPTGTQPT-----TTTWTPTPTPTGTQPT- 163  
QY 510 APSTVVPAPPTAPPAAPVPTPTTAIPTAPPAAPTAAAPTAAAPTAAAPTAAAPT 569  
Db 163 -----PTTPTITTTTWTPT-----PTPTGTPTPTPTPT-----TTTWTPTPTPTSTQ 207  
QY 570 TTAVPEIPITVTSAPTAAPTAAAPTAAAPTAAVPEIPT-TVTSPTAAAPTAAAPT 628  
Db 208 STPTPTTTTWTPTPTPTGTQPTPTPTTTTWTPTPTPTPTPTPTPTPTPTPT 262  
QY 629 TTVTPTTAAPTAAAPNTTVAAPTAAAPTAAAPTAAAPNTTVAAPTPT 673  
Db 263 TTVTPT-----PTPTSTQRTPTPTIT-----TTTWTPTPT 294  
RESULT 13  
A38420  
C:Species: glycoprotein precursor - black rockcod  
C:Species: Notothenia coriiceps neglecta (black rockcod)  
C:Date: 28-Jun-1991 #sequence\_revision 30-Jan-1993 #text\_change 09-Sep-1997  
C:Accession: A38420  
R:Hsiao, K.C.; Cheng, C.H.C.; Fernandes, I.E.; Detrich, H.W.; DeVries, A.L.  
Proc. Natl. Acad. Sci. U.S.A. 87, 9265-9269, 1990  
A:Title: An antifreeze glycopeptide gene from the antarctic cod Notothenia coriiceps  
A:Reference number: A38420; MUID:91067687  
A:Accession: A38420  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-822 <HSI>  
A:Cross-references: GB:M55000; NID:g213391; PID:g213392  
A:Note: the authors translated the codon CCT for residue 489 as Ala  
C:Keywords: glycoprotein

Query Match 8.5%; Score 374; DB 2; Length 822;  
Best Local Similarity 25.7%; Pred. No. 6e-12;  
Matches 193; Conservative 21; Mismatches 332; Indels 204; Gaps 26;

QY 65 APRDCAPGTEFKESAQTCVHAA-----LAGCTLPQPP-----AETQAPATTQAPT 111  
Db 146 AATPATPAFHFAATAATPATPALIFAATAATAATPATPATPAFHFAATAATPATPAL 205  
QY 112 TQAPTITQOA-PTTITQAPTITQAPTITQAPTITQAPTITQAPTITQAPTIT 170  
Db 206 IFATAATAATPATPALIFAATAATAATPATPALIFAATAATAATPATPALIFAATA 265  
QY 171 TQAPTITQAPTITQAPTITQAPTITQOA-----ATTPA-----ATTPA-ATTP- 214  
Db 266 ATPATAACNFAATAATPATPALIFAATAATAATPATPALIFAATAATAATPATPALI 325  
QY 214 -AATTPAATTPA-----ATTPGVAPTSAPVWPPICELLNPGCPADFIDHLLPHDKYC 266  
Db 326 FAATAATAATPATPALIFAATAATAATPATPALIFAATAATAATPATPALIFAATA 357  
QY 267 NLFYQCSNGYFEQRCPEGLYFNYPVQRCDSANVEDGEISAPPVTEGNEDEDIDGD 326  
Db 357 -----ALIFAATAATAATPATPALIFAATAATAATPATPALIFAATA 397  
QY 327 LLDNGCPANFEIDWLLPHGNRCDKYYQVHGNLVERRCGAGT-----HFSFELQOCDHIEL 382  
Db 398 AATAATPATPALIFAATAATAATAATAATAATAATAATAATAATAATAATAATA 430  
QY 383 VGTCLPGSESEVDVDEDACTGWCPTPIEWEPLNCGPADFSIDHLLPHESDCGQYLO 442  
Db 431 TAATAATAATPATPALIFAATAATAATAATAATAATAATAATAATAATAATAATA 479  
QY 443 CVHGQTIARPCPNLHFSFATOSCEPVYTAGQVFECDSDNOCTSTAAPTAAAPTAA 502  
Db 480 AATAATAATPATPALIFAATAATAATAATAATAATAATAATAATAATAATAATA 531

QY 503 PTAAPTAAABST-----VVPATPATATA-----APVPTTAITPT--APTAAPTAA 546  
Db 532 ATAATAATATAAFNAFAATAATAATATAALNFAATAATAATATAATPALIFAATAATAATP 591  
QY 547 TTAAPESPVTTPVPTTAAPTA-----APTTAVPEIP----- 578  
Db 592 ATAALNFAATAATATAATPALIFAATAATAATPATPAFHFAATAATATAATPATPALIFAA 651  
QY 578 --ITVTSAPTAAPTAAPTAAPTAVEIPTTTVSPTAA--PTAAPAPNTTIVTVP 634  
Db 652 TAATAATATAALNFAATAATAATATAALNFAATAATAATATAATPALIFAATAATAATPA 711  
QY 635 TAA-----PTTAA--PAPNTTIVTAPTAAPTAA--PAPNTTIVTVPPTAAPT---- 678  
Db 712 TAALNFAATAATAATATAATPALIFAATAATAATATAAFNAFAATAATAATATAATPALIF 771  
QY 678 AAPPTVAHAPNTTAAPV---TTTSPATTP 704  
Db 772 AATAATAATPATPATPALIFAATAATAATP 801  
RESULT 14  
A:48292  
mucin, tracheobronchial - dog  
C:Species: Canis lupus familiaris (dog)  
C:Date: 27-Jun-1994 #sequence\_revision 27-Jun-1994 #text\_change 10-Sep-1997  
C:Accession: A48292  
R:Verma, M.; Davidson, E.A.  
Proc. Natl. Acad. Sci. U.S.A. 90, 7144-7148, 1993  
A:Title: Molecular cloning and sequencing of a canine tracheobronchial mucin cDNA contain  
A:Reference number: A48292  
A:Accession: A48292  
A>Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-1118 <VER>  
A:Cross-references: GB:X69164; NID:9402557; PID:9402558  
A:Note: authors translated the codon CAA for residue 52 as Gly, GAT for residue 94 as As  
C:Comment: It is uncertain whether Met-1 or Met-52 is the initiator.  
C:Keywords: glycoprotein

Query Match 8.5%; Score 373; DB 2; Length 1118;  
Best Local Similarity 23.5%; Pred. No. 8.7e-12;  
Matches 209; Conservative 52; Mismatches 309; Indels 320; Gaps 37;  
QY 66 PRDCAPGTEFKSAOTCVHAALAGCTLP-----GPPAETQAPATTQAPTTQAP 115  
Db 203 PQCLPTPTPLSTPSIIRPOEALPLCLSAVSETPTPISTTTTPTPTPTGTQTP 262  
QY 116 TTTTQAPTTTQAPTTT--TQAPT----- 140  
Db 263 TTTPTTTTPTPTPTGTGTPTTTLTTTMTPTPTSQKLGSDDTPVLVIALVTPT 322  
QY 141 ARTTQAPTTTQAPTTTQAPTTT---TQAPTTTQAPTTTQAPT----- 183  
Db 323 TPTGTPTPTTTTPTTTTPTPTGTGTPTTPTTPTTPTTPTTSTKVLVTPXTT 382  
QY 183 -----TTQAPTTTQAPTTTQAPTTTQAPTTAATPA-----ATTAA 220  
Db 383 TTGVTPTTHRHTRHQTTHHTTHTTHTTATATGTGRTPTTTLTHTTQTTTHRQTRPQP 442  
QY 221 TTPAATTPGVPAPTAPVMPVICELPNCPADFDHLLPHDKYCNLFYQCSNGYTFFQ 280  
Db 443 TTPSPTTTIVD--PTQHPSPPLGDRPN--THRDTPN-----HDTTIT--TWTPTPTPR 494  
QY 281 RCPGLYFNPNYVORCDSPANVECDGEISAPPVTEGNE-----BDI 322  
Db 495 TRPQPRYTTHTTMTPT-----TPTPTHTDHHDLPPPPQVRHRTPRRCHL 545  
QY 323 DIGDLL--DNGCPANFEIDWLLPHGNRCDKYQCVH-----GNLVERRCAG 367  
Db 546 TSGSPLTHYPNTTNTN-----LLEKNRSELINTPTHTMIISAAAMATPPSLVHRCRSG 599

QY 368 THFSPELQQ-----CDHIEL-----VGCTLPGG-----ESEVDVDEAC 402  
Db 600 -HSSRAQKSRASGCLNMLGVESGVGVSTAGGALVRSRCRTRHSRAEE----- 651  
QY 403 TGWYCPTPEIEWEPLNCGCPADFSIDHLLPHES-----DCGQVLCQVHGOTIARPCPG 455  
Db 651 -----GLGPEAVVSMMLGVESGVGVGRHCGGRHLHSAGGG--AEDDPG 691  
QY 456 NL-----HFSPATQCESPV-----TAGCQVFECDSNDOCTSTAAPTA 493  
Db 692 RVPRVGTALDSTQHTSRRAICKAGVPNEELTYTACASALQSETCADMIATATIHHTT 751  
QY 494 APT-----AAPTAAPTAAAPTAAPTVV-----PPATPPATAAPVPT 530  
Db 752 SRSIRKPKTISLATTYACASTAASVGTVQLDMATGGLRDRGMAGAMPAGDSRGSPCT 811  
QY 531 TAITPTPATTAAPTAAAPESPPTTIVTVPPTTAAPTAAPTTAVPEIPTVTSAPTAAPT 590  
Db 812 ELEFRPGLQAPTPTPIS-----TTTTTPTPTPTG-----TOTPTTTTSP 852  
QY 591 APTAAPTAAPTT--AVPEIPTTVPSPPTAAPT----- 622  
Db 853 PPPGDPKTGKTVYKIIGSLRTTFNLPAWLGYYITICKLQMSNRKTKTAETINVKTLGSLM 912  
QY 622 --AAPAPNTTIVTP--PTAAPTAAAPTAAPTNT-----VTAPPT--AAPTTAA 661  
Db 913 MDVTPINSCRVTTPPPSGKPCADSRMLNPTAETGSGVIGAGCCCVLTPTTTHRVIRCTA 972  
QY 662 PAPNTTIVTVPPTAAPTAAPTTAVHAPNTTAAPTVTTSPATPTPEDDDIDP 711  
Db 973 TPT-TPISTTTTPTPTQHP--AHRP--TTSPTITTTT-VTQPDTHTDP 1016  
RESULT 15  
SL5921  
Protein TPX-VT3 - Thermoproteus phase 1 (strain strain KRA1 10/12)  
C:Species: Thermoproteus phase 1, TTV1  
A:Variety: strain KRA1 10/12  
C:Date: 21-Nov-1993 #sequence\_revision 26-May-1995 #text\_change 09-Sep-1997  
C:Accession: SL5921; S06686  
R:Neumann, H.; Zillig, W.  
Nucleic Acids Res. 18, 2171, 1990  
A:Title: Nucleotide sequence of the viral protein TPX of the TTV1 variant VT3.  
A:Reference number: SL5921; MUID:90245666  
A:Accession: SL5921  
A:Molecule type: DNA  
A:Residues: 1-474 <NEW>  
A:Cross-references: EMBL:X14717; NID:962191; PID:e20711; PID:g584349  
A:Experimental source: strain KRA1 10/12  
A:Note: the authors translated the initiation codon GTG for residue 1 as Val  
R:Neumann, H.  
submitted to the EMBL Data Library, March 1989  
A:Reference number: S06686  
A:Accession: S06686  
A:Molecule type: DNA  
A:Residues: 207-474 <NEW>  
A:Cross-references: EMBL:X14717  
A:Experimental source: strain KRA1 10/12  
C:Genetics:  
A:Start codon: GTG  
Query Match 8.4%; Score 367.5; DB 2; Length 474;  
Best Local Similarity 32.6%; Pred. No. 8e-12;  
Matches 107; Conservative 20; Mismatches 132; Indels 69; Gaps 13;  
QY 394 EYVDDEACTGMYCPTPEIEWEPLNCGPADFSIDHLLPHESDCGYLQCVHGOTIARPC 453  
Db 192 EIDISQFAF-----TDLVMVLSINGSTANVKV-----YKQ---GSNI----- 228  
QY 454 PGNLHFSPATQSCSPVTAQCO-----VFECDSNDOCTSTAAPTAAPTAAPTA 503  
Db 228 -GTVSGNYSSTPYGNPSMAGYGTVDKHYANFVLFPEDPQVTVTPISPSPTPTPTP 286



Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	670	15.2	5179	1	MUC2_HUMAN	Q02817 homo sapien
2	457	10.4	865	1	CPN_DROME	Q02910 drosophila
3	446.5	10.2	797	1	VLGX_HSVB	P28968 equine herp
4	430.5	9.8	662	1	MUC1_XENLA	Q05049 xenopus lae
5	428	9.7	1255	1	MUC1_HUMAN	P15941 h mucin 1 p
6	410	9.3	1367	1	AMTH_YEAST	Q08640 saccharomyc
7	372.15	8.5	1664	1	SLP1_CLOTH	Q06852 clostridium
8	369.5	8.4	822	1	ANP_NOTCO	P24856 notothenia
9	367.5	8.4	474	1	VP3_TTV1V	P19275 thermoprote
10	357	8.1	279	1	Y091_NPVOP	O10341 orgyia pseu
11	352	8.0	670	1	VG50_HSV1Z	Q00130 ictalurid h
12	336.5	7.7	267	1	EXTN_WAIZE	P14918 zea mays (m
13	331	7.5	400	1	MUAI_XENLA	P10667 xenopus lae
14	331	7.3	907	1	VG3_EBV	Q03200 Epstein-bar
15	319.5	7.3	860	1	CH12_COCIM	P03200 Epstein-bar
16	317	7.2	307	1	SGS3_DROME	P54197 coccidioid
17	315	7.2	1161	1	YJ99_YEAST	P02840 drosophila
18	313.5	7.1	886	1	VG33_EBVAB	P47179 saccharomyc
19	311	7.1	620	1	EXTN_TOBAC	Q07284 Epstein-bar
20	307.5	7.0	283	1	EXTN_SORVU	P13983 nicotiana t
21	300.5	6.8	3149	1	TEGU_EBV	P24152 sorghum vul
22	296.5	6.7	634	1	HW11_CANAL	P23186 Epstein-bar
23	290.5	6.6	1794	1	YAV1_SCHPO	P46593 candida alb
24	290	6.6	1162	1	TCNA_TRYCR	Q10172 schizosacch
25	285.5	6.5	1251	1	YQ23_CAEEL	P23253 trypanosoma
26	283.5	6.5	2476	1	ZAN_PTG	Q09550 caenorhabdi
27	282	6.4	699	1	VGJG_HSV2H	Q28983 sus scrofa
28	276.5	6.3	338	1	SEDI_YEAST	P13290 herpes simp
29	270.5	6.2	263	1	SGS3_DROYA	Q01589 saccharomyc
30	269.5	6.1	1083	1	T2D3_HUMAN	P13728 drosophila
31	268.5	6.1	796	1	YS9A_CAEEL	Q00268 homo sapien
32	268	6.1	3421	1	TEGU_HSVCD	Q09625 caenorhabdi
33	267.5	6.1	600	1	SP96_DICDI	P28955 equine herp
34	266	6.1	721	1	YV21_MYCTU	P14328 dictyosteli
35	265.5	6.0	901	1	A180_MOUSE	Q10690 mycobacteri
36	261.5	5.9	467	1	RPB1_CRIGR	Q61548 mus musculu
37	261.5	5.9	1970	1	RPB1_HUMAN	P11414 cricetulus
38	261.5	5.9	1970	1	RPB1_MOUSE	P24928 homo sapien
39	261.5	5.9	439	1	XP2_XENLA	Q08775 mus musculu
40	260	5.9	915	1	A190_RAT	P17437 xenopus lae
41	253.5	5.8	1802	1	HKR1_YEAST	Q05140 rattus norv
42	252.5	5.7	1742	1	GUNA_CALSA	P41809 saccharomyc
43	250.5	5.7	817	1	VRP1_YEAST	P22534 caldocellum

```

RESULT      1
ID          MUC2_HUMAN    STANDARD;   PRT;  5179 AA.
AC Q02817; Q14878;
DT 01-JUN-1994 (REL. 29, CREATED)
DD 01-NOV-1997 (REL. 35, LAST SEQUENCE UPDATE)
DE 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
GN MUCIN 2 PRECURSOR (INTESTINAL MUCIN 2).
OS MUC2 OR SMUC.
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA; EUTHERIA; PRIMATES.
[1]
RN RP     SEQUENCE FROM N.A.
RC TISSUE=INTESTINE;
RX MEDLINE; 94132002.
RL GUM J.R., JR.; HICKS J.W., TORIBARA N.W., SIDDIKI B., KIM Y.S.;
RNL [J.] J. BIOL. CHEM. 269:2440-2446(1994).
RP RC     SEQUENCE OF 636-1895 AND 4196-5179 FROM N.A.
TX TISSUE-COLON;
RY MEDLINE; 93016075.
RA GUM J.R., JR.; HICKS J.W., TORIBARA N.W., ROTHE E.-M., LAGACE R.E., KIM Y.S.;
RAL [BIO]. CHEM. 267:21375-21383(1992).
RN RP     SEQUENCE OF 1343-1895 AND 4176-4195 FROM N.A.
RX MEDLINE; 91358717.
RR TORIBARA N.W., GUM J.R. JR., CULHANE P.J., LAGACE R.E., HICKS J.W., PETERSEN G.M., KIM Y.S.;
RJ CLIN. INVEST. 88:1005-1013(1991).
CC -! FUNCTION COATS THE EPITHELIAL CELLS OF THE INTESTINES, AIRWAYS, AND OTHER MUCUS MEMBRANES CONTAINING GLYCOPROTEINS.
CC -! SUBUNIT MULTIMERIC.
CC -! SURCELLULAR LOCATION SECRETED.
CC -! TISSUE SPECIFICITY COLON, SMALL INTESSTINE, COLONIC TUMORS, BRONCHUS, CERVIIX AND GALL BLADDER.
CC -! PTM ALL CYSTEINE RESIDUES ARE INVOLVED IN INTRACHAIN OR INTERCHAIN DISULFIDE BONDS (BY SIMILARITY).
CC -! POLYMORPHISM THE NUMBER OF REPEATS IS HIGHLY POLYMERPHIC AND VARIES AMONG DIFFERENT ALLELES.
CC -! SIMILARITY THE N-TERMNAL DOMAIN SHOWS SOME SIMILARITY TO THAT OF SILKWORM HEMOCYTIN.
CC -! SIMILARITY CONTAINS 2 WFEC DOMAINS.
CC -! SIMILARITY CONTAINS 1 C-TERMINAL CYSTINE KNOT-LIKE DOMAIN (CTCK) EMBL; I21998; G45A1154; -.
DR DR EMBL; W74027; GI88864; -.
DR EMBL; M94131; GI86396; -.
DR EMBL; M94132; GI86398; -.
DR HSP; PF5445; IPSH.
DR MM; F156370; -.
DR PROSITE; PS00022; EGF_1; UNKNOWN_1.
DR PROSITE; PS01185; CTCK_1; 1.
DR PROSITE; PS01225; CTCK_2; 1.
DR PROSITE; PS01208; WFEC_2.
RW GLYCOPROTEIN; REPEAT; POLYMORPHISM; SIGNAL.
FT SIGNAL      1       20
FT CHAIN      21    5179
FT DOMAIN     1401    1747 APPROXIMATE REPEATS.
FT REPEAT     1417    1416        1.
FT REPEAT     1417    1432        2.
FT REPEAT     1433    1448        3.
FT REPEAT     1449    1464        4.
FT REPEAT     1465    1471        5.
FT REPEAT     1472    1478        6.
FT REPEAT     1479    1494        7A.

```

```
FT REPEAT 1495 1517 7B.
FT REPEAT 1518 1533 8A.
FT REPEAT 1534 1556 8B.
FT REPEAT 1557 1572 9A.
FT REPEAT 1573 1596 9B.
FT REPEAT 1597 1612 10A.
FT REPEAT 1613 1635 10B.
FT REPEAT 1636 1651 11A.
FT REPEAT 1652 1675 11B.
FT REPEAT 1676 1699 12.
FT REPEAT 1699 1715 13.
FT REPEAT 1700 1715 14.
FT REPEAT 1716 1731 15.
FT REPEAT 1732 1747 16.
FT DOMAIN 4815 4886 WNF 1.
FT DOMAIN 4924 4991 WNF 2.
FT DOMAIN 5075 5160 CTCK.
FT DISULFID 5075 5126 BY SIMILARITY.
FT DISULFID 5089 5136 BY SIMILARITY.
FT DISULFID 5088 5152 BY SIMILARITY.
FT DISULFID 5102 5154 BY SIMILARITY.
FT CARBOHYD 163 163 POTENTIAL.
FT CARBOHYD 423 423 POTENTIAL.
FT CARBOHYD 670 670 POTENTIAL.
FT CARBOHYD 770 770 POTENTIAL.
FT CARBOHYD 894 894 POTENTIAL.
FT CARBOHYD 1139 1139 POTENTIAL.
FT CARBOHYD 1154 1154 POTENTIAL.
FT CARBOHYD 1215 1215 POTENTIAL.
FT CARBOHYD 1230 1230 POTENTIAL.
FT CARBOHYD 1246 1246 POTENTIAL.
FT CARBOHYD 1787 1787 POTENTIAL.
FT CARBOHYD 1820 1820 POTENTIAL.
FT CARBOHYD 4339 4339 POTENTIAL.
FT CARBOHYD 4351 4351 POTENTIAL.
FT CARBOHYD 4362 4362 POTENTIAL.
FT CARBOHYD 4373 4373 POTENTIAL.
FT CARBOHYD 4422 4422 POTENTIAL.
FT CARBOHYD 4438 4438 POTENTIAL.
FT CARBOHYD 4502 4502 POTENTIAL.
FT CARBOHYD 4616 4616 POTENTIAL.
FT CARBOHYD 4627 4627 POTENTIAL.
FT CARBOHYD 4752 4752 POTENTIAL.
FT CARBOHYD 4787 4787 POTENTIAL.
FT CARBOHYD 4881 4881 POTENTIAL.
FT CARBOHYD 4888 4888 POTENTIAL.
FT CARBOHYD 4955 4955 POTENTIAL.
FT CARBOHYD 4970 4970 POTENTIAL.
FT CARBOHYD 5019 5019 POTENTIAL.
FT CARBOHYD 5038 5038 POTENTIAL.
FT CARBOHYD 5089 5089 POTENTIAL.
FT CONFLICT 1351 1351 H -> L (IN REF. 3).
FT CONFLICT 1412 1412 T -> S (IN REF. 3).
FT CONFLICT 1449 1449 L -> P (IN REF. 3).
FT CONFLICT 1504 1504 M -> T (IN REF. 3).
FT CONFLICT 4192 4192 G -> S (IN REF. 2).
SQ SEQUENCE 5179 AA; 540295 MW; ECD04582 CRC32;

Query Match 15.2%; Score 670; DB 1; Length 5179;
Best Local Similarity 30.6%; Pred. No. 7.9e-22;
Matches 220; Conservative 42; Mismatches 262; Indels 196; Gaps 31;

QY 91 TLPGPP-----AETQAPATTQAP-----TTTQAPTTTQAPTTTQAP-----TTTQ 134
D 1537 TTPSPPTTPIPTTSTTLPLPTTSPPTTTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1596
QY 135 APTTQAP-----TTTQAPTTTQAPTTT-----QAPTTTQAPTTTQAP-----TTTQ 184
D 1597 PPTTSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTP 1656
QY 185 QAPTTTQAPT-TITQAAATTAATPAA-----TTTAA-----TTTAA 219
```

```
Db 1657 PSPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1716
QY 220 ATT---PAATTPGVPAPTSAP-----VWPP-----ICELLPGCPADFIDHLI 260
Db 1717 STTPSPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTTTPPTT 1772
QY 261 PHDKYCNLFYQCSNGYTFEQRCPEGLYFNXYVQRCDSPANVECDGEISPAFPVTEGNEDE 320
Db 1772 -----PESTTTPT-----TPCVPLCNMTGWL-----SGKPNFHKPGGDT 1806
QY 321 DIDIGDLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRCGAGTHFSEFLOQCCHI 380
Db 1807 EL-IGDVGCGPWAANISC-----RATMYPDVPIGOL----- 1837
QY 381 ELVGCTLPGGESEVDDVEDACTGWYCPTETIEWEPLNG-CPADFSIDHLLPHESDCGQ 439
Db 1837 -----GQTVVCDVS-----VGLCKNE-----DQKGGVLPKAFCLNVEI----- 1872
QY 440 YLQCVHGOTIARPC-----PCNLHFSPATQCESPVTAGCOVFECDSNQCTS 487
Db 1872 NVQCC--ECVTQPTTMTTNTTENPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1926
QY 488 TAAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 544
Db 1927 TTTVTPTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 1986
QY 545 APTTAAPESPVTTPVPTTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 604
Db 1987 TPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2046
QY 605 PEIPTTVTSPPTAAP-----TTAAPAPNTTVVPTAAP-----TTAAPAPNTTVIAPPTAA 656
Db 2047 PTPTPTGQTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2106
QY 657 P-----TTAAPAPNTTVVPTAAP-----TAAP-----PVAHAPNTTAAPVTTTSAPATTP 704
Db 2107 PTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 2166

RESULT 2
CPN_DROME CPN_DROME STANDARD; PRT; 865 AA.
AC 002910;
DT 01-OCT-1993 (REL. 27, CREATED)
DT 01-OCT-1993 (REL. 27, LAST SEQUENCE UPDATE)
DE CALPHOTIN.
GN CPN OR CAP.
OS DROSOPHILA MELANOGASTER (FRUIT FLY).
OC EUKARYOTA; METAZOA; ARTHROPODA; INSECTA; DIPTERA.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE; 93165730.
RA BALLINGER D.G., XUE N., HARSHMAN K.D.;
RA MARTIN J.H., BENZER S., RUDNICKA M., MILLER C.A.;
RL PROC. NATL. ACAD. SCI. U.S.A. 90:1531-1535(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE; 93165730.
RA BALLINGER D.G., XUE N., HARSHMAN K.D.;
RL PROC. NATL. ACAD. SCI. U.S.A. 90:1536-1540(1993).
CC -!- FUNCTION: MIGHT FUNCTION AS A CALCIUM-SEQUESTERING "SPONGE" TO
CC OF CA+2 PER MOL OF PROTEIN.
CC -!- SUBUNIT: HOMODIMER (PROBABLE).
CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC; HYPODENSE COMPARTMENT.
CC -!- TISSUE SPECIFICITY: SOMA AND AXONS OF PHOTORECEPTOR CELLS OF
CC COMPOUND EYES AND OCELLI.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED EARLY IN PHOTORECEPTOR CELL
CC DEVELOPMENT.
DR EMBL; L02111; G157032; -.
```





QY 521 PATAAPVPPTTAPPTAAPTAAAPESPTTIVPPTAAPTAAAPTAVPEIPTV 580  
DB 410 -----TTTTTTTTTKATTTTTTTTTT-----PTTTT----- 439  
QY 581 TSAPTAAPTAAPTAAAPTAAAPTAVPEIPTTSPPTAAPTAAAPTAVPPTAAPT 640  
DB 439 -----TTTTTKATTTPTTTP-----TTTTTKATTTPTTPTTTTATTTPTT 488  
QY 641 AAPAPNTVTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 700  
DB 489 TTTTPTTTTATTTTPTT-----TTTTT-----TTTKATTTTSGE 525  
QY 701 ATTPEDDDIDPLNDPINC-----VEECNVLPAHADC--- 736  
DB 526 CKMEPSKRADCGPGITESQCKSGCCFSSIPQTKWCFYSLPQVADCKVAPSSRVDCGF 585  
QY 736 -----DRYWCVDGNNQVLLVWCSEGLQFNPTTKTCD-----F 766  
DB 586 GGITADOCQRNCCFSSISGFKWCFYSTSQGNAMCS-----GPTKRRCDCGYPGSISSV 640  
QY 767 ACNVCGRSNIQMSSEYEGVQVFIW 792  
DB 641 CINRGCCWDSVMN-----VPW 657  
RESULT 5  
ID MUC1\_HUMAN STANDARD; PRT: 1255 AA.  
AC P15941: P15942: P13931: P17626;  
DT 01-JAN-1980 (REL. 13, CREATED)  
DT 01-APR-1990 (REL. 14, LAST SEQUENCE UPDATE)  
DE MUCIN 1 PRECURSOR (POLYMORPHIC EPITHELIAL MUCIN) (PEM) (PEMT)  
DE (EPISIALIN) (TUMOR-ASSOCIATED MUCIN) (CARCINOMA-ASSOCIATED MUCIN)  
DE (TUMOR-ASSOCIATED EPITHELIAL MEMBRANE ANTIGEN) (EMA) (H23AG)  
DE (PEANUT-REACTIVE URINARY MUCIN) (PUM).  
GN MUC1.  
OS HOMO SAPIENS (HUMAN).  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; TETRAPODA; MAMMALIA;  
OC EUTHERIA; PRIMATES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=PANCREAS;  
RX MEDLINE: 90368716.  
RA LAN M.S., BATRA S.K., QI W.-N., METZGAR R.S., HOLLINGSWORTH M.A.;  
RL J. BIOL. CHEM. 265:15294-15299(1990).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 90202794.  
RA LIGTENBERG M.J.L.; VOS H.L.; GENNISSSEN A.M.C.; HILKENS J.;  
RL J. BIOL. CHEM. 265:5573-5578(1990).  
RN [3]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BREAST CARCINOMA;  
RX MEDLINE: 90368715.  
RA GENDLER S.J., LANCASTER C.A., TAYLOR-PAPADIMITRIOU J., DUHIG T.,  
RA PEAT N., BURCHELL J., PEMBERTON L., LALANI E.-N., WILSON D.;  
RL J. BIOL. CHEM. 265:15286-15293(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91097524.  
RA LANCASTER C.A., PEAT N., DUHIG T., WILSON D.,  
RA TAYLOR-PAPADIMITRIOU J., GENDLER S.J.;  
RL BIOCHEM. BIOPHYS. RES. COMMUN. 173:1019-1029(1990).  
RN [5]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BREAST CARCINOMA;  
RX MEDLINE: 90276413.  
RA WRESCHNER D.H., HAREUVENI M., TSAREFATY I., SMORODINSKY N.,  
RA HOROV J., ZARETSKY J., KOTKES P., WEISS M., LATHE R., DION A.,  
RA KEYDAR I.;  
RL EUR. J. BIOCHEM. 189:463-473(1990).

RN [6]  
RP SEQUENCE FROM N.A.  
RC TISSUE-BREAST CARCINOMA;  
RX MEDLINE: 90276414.  
RA HAREUVENI M., TSAREFATY I., ZARETSKY J., KOTKES P., HOREV J.,  
RA ZRIHAN S., WEISS M., GREEN S., LATHE R., KEYDAR I., WRESCHNER D.H.;  
RL EUR. J. BIOCHEM. 189:475-486(1990).  
RN [7]  
RP SEQUENCE FROM N.A.  
RX MEDLINE: 91033045.  
RA TSAREFATY I., HAREUVENI M., HOREV J., ZARETSKY J., WEISS M.,  
RA JELTSCH J.M., GARNIER J.M., LATHE R., KEYDAR I., WRESCHNER D.H.;  
RL GENE 93:313-318(1990).  
RN [8]  
RP PARTIAL SEQUENCE FROM N.A.  
RX MEDLINE: 88330762.  
RA GENDLER S.J., TAYLOR-PAPADIMITRIOU J., DUHIG T., ROTHARD J.,  
RA BURCHELL J.;  
RL J. BIOL. CHEM. 263:12820-12823(1988).  
RN [9]  
RP SEQUENCE OF 1-46 FROM N.A.  
RC TISSUE-BREAST CARCINOMA;  
RA BULUWELA L., LIU Q., LUOMANI Y.A., GOMM J.J., COOMBS R.C.;  
RL SUBMITTED (OCT-1992) TO EMBL/GENBANK/DBJ DATA BANKS.  
CC -!- FUNCTION: DIRECT OR INDIRECT INTERACTION WITH ACTIN  
CYTOSKELETON.  
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. A SECRETED FORM  
IS ALSO PRODUCED.  
CC -!- TISSUE SPECIFICITY: ABERRANTLY EXPRESSED IN HUMAN EPITHELIAL  
TUMORS, SUCH AS BREAST CANCER.  
CC -!- PTM: HIGHLY GLYCOSYLATED (N- & O-LINKED CARBOHYDRATES & SIALIC  
ACID).  
CC -!- POLYMORPHISM: THE NUMBER OF REPEATS IS HIGHLY POLYMORPHIC AND  
VARIES AMONG DIFFERENT ALLELES. THE BREAST TUMOR MUC1 HAS  
GENERALLY ONLY 5 TO 7 REPEATS.  
CC -!- ALTERNATIVE PRODUCTS: VARIOUS VARIANTS ARE PRODUCED BY ALTERNATIVE  
SPLICING OF THE MUC1 GENE.  
DR EMBL: J05582; G189599; -.  
DR EMBL: M32738; G182124; -.  
DR EMBL: M32739; G182129; -.  
DR EMBL: J05581; G188870; -.  
DR EMBL: M61170; G189784; -.  
DR EMBL: X52229; G37054; ALT\_SEQ.  
DR EMBL: X52228; G36435; ALT\_SEQ.  
DR EMBL: M35093; G182253; ALT\_SEQ.  
DR EMBL: Z17324; G34807; -.  
DR EMBL: Z17325; G34809; -.  
DR EMBL: M21868; G386963; ALT\_SEQ.  
DR PIR: A35175; A35175.  
DR PIR: B35175; B35175.  
DR PIR: S10218; S10218.  
DR MIM: 158340; -.  
KW GLYCOPROTEIN; SIGNAL; CYTOSKELETON; ACTIN-BINDING; TRANSMEMBRANE;  
KW REPEAT; ALTERNATIVE SPLICING; POLYMORPHISM.  
FT SIGNAL 1 23  
FT CHAIN 24 1255  
FT DOMAIN 24 1162 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 1163 1186 POTENTIAL.  
FT DOMAIN 1187 1255 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 81 960 44 X 20 AA TANDEM REPEATS.  
FT CARBOHYD 957 957 POTENTIAL.  
FT CARBOHYD 975 975 POTENTIAL.  
FT CARBOHYD 1029 1029 POTENTIAL.  
FT CARBOHYD 1055 1055 POTENTIAL.  
FT CARBOHYD 1133 1133 POTENTIAL.  
FT VARSPPLIC 19 19 T -> ATAPKPAT (IN FORM B).  
FT VARSPPLIC 20 22 MISSING (IN FORM C).  
FT VARSPPLIC 20 31 MISSING (IN FORM D).  
FT VARSPPLIC 126 905 MISSING (IN POLYMORPHIC EPITHELIAL  
VARIANT).  
FT VARSPPLIC 1077 1087 FLQYKGGFL -> VSIGLSFPMPLP (IN SECRETED  
FORM).  
FT VARSPPLIC 1088 1255 MISSING (IN SECRETED FORM).



FT CONFLICT 1021 1021 S -> T (IN REF. 3).  
FT CONFLICT 1251 1251 A -> T (IN REF. 3).  
SQ SEQUENCE 1255 AA; 122072 MW; 806B1004 CRC32;

Query Match 9.7%; Score 428; DB 1; Length 1255;  
Best Local Similarity 26.7%; Pred. No. 4.8e-12;  
Matches 208; Conservative 45; Mismatches 294; Indels 232; Gaps 37;

QY 17 RPEVSDAEKNPALHE--PHDPXPPAEQXXLLPXEDCTFYCEYGLKFIAPRDCAPGTE 74  
DB 105 RPAIGST--TPPAHDVTSAPDNKPAGSTAPPA-----HGVTSAPDTRPAGS- 151  
QY 75 FKFSATCVHAALAGCTLPGPPAEITTOAATTOAQTTOAQTTOAQTTOAQTTOAQTTOA 127  
DB 151 ---TAPPAHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPA 207  
QY 128 APTTTTQAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOA 186  
DB 208 PGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPA 267  
QY 187 PTTTQAQT--TITQAAT--PAATTAA--TTPAATTAAATPAATTCVPAPTSAP-- 238  
DB 268 PGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPA---PGSTAPPAGHGVTSAPDT 324  
QY 238 -----VWPPICELLNGCPADFIDHLLPHDKYCNLFYQCSNGYTF--EQRCPEGLYFN 289  
DB 325 RPAGGSTAPP-----AHG-----AHGVTSAPTDRPAGSTAP 353  
QY 290 PYVQRCDSANVECDGEISPAVPVTEGNEDEDIDGLLDNGCPANFEIDWLLPHGNRCD 349  
DB 354 PAHGVTSAPDTRPAG--STAPP----- 375  
QY 350 KYQCVHGNLVERRCGAGTHFSFELQOCDHIELVGCPLPGGESEEDVDEDACTGWYCP- 409  
DB 375 -----AHG-----VTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPP 414  
QY 409 -----TEPIEWELPNGCPADFSIDHLLPHESDCGOYLQCVHGTIA---RPCGNLHFGP 461  
DB 415 AHGVTSAPTDRPAP-----GSTAPPAGHGVTSAPDTRPAGSTAPPAGST--APP 454  
QY 462 ATQSCSPYTAGQVFCDSNOCSTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 519  
DB 455 AHGVTSA-----DTRPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDT 504  
QY 520 PPATAAPVPP---TTAIPTPAATAAPTAAAPESPTTVTPPTAAAPTAAAPTAAVPEI 576  
DB 505 RPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDT 564  
QY 577 -PTVTSAP-----TAAPTAAPTAAPTAAPTAAVPEIPTT-----VTSPP-----TAAPT 621  
DB 565 RPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDT 624  
QY 622 AAPANTVTVPP-----TAAPTAAAPNTVTVAPP-----TAAPTAAAPNTVTVPT 673  
DB 625 -RPAGS--TAPPAGHGVTSAPDT--RPAGS--TAPPAGHGVTSAPDT--RPAGS--TAPPA 675  
QY 674 AAPTAAPT-----VAHAPNTAAAPVTT-----TSAPATTPEDDDIDPP 712  
DB 676 HGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPPAGHGVTSAPDTRPAGSTAPP 734

## RESULT 6

AMYL YEAST  
ID AMYL YEAST STANDARD; PRT: 1367 AA.  
AC P08640; P08068;  
DT 01-AUG-1988 (REL. 08, CREATED)  
DT 01-FEB-1995 (REL. 31, LAST SEQUENCE UPDATE)  
DT 15-JUL-1998 (REL. 36, LAST ANNOTATION UPDATE)  
DE GLUCOAMYLASE S1/S2 PRECURSOR (EC 3.2.1.3) (GLUCAN 1,4-ALPHA-  
DE GLUCOSIDASE) (1,4-ALPHA-D-GLUCAN GLUCOHYDROLASE).  
GN STAL OR ST42 OR ST45 OR YI019C.  
OS SACCHAROMYCES CEREVISIAE (BAKER'S YEAST).

OC  
RN  
RP  
RC  
RA  
RA  
RA  
RA  
RA  
RA  
RA  
RL  
RL  
RN  
RP  
RC  
RX  
RA  
RL  
RN  
RP  
RC  
RX  
RA  
RL  
RN  
RP  
CC  
CC  
DR  
DR  
DR  
DR  
DR  
DR  
DR  
DR  
DR  
DR  
KW  
KW  
FT  
FT  
FT  
FT  
SQ  
Query Match 9.3%; Score 410; DB 1; Length 1367;  
Best Local Similarity 19.9%; Pred. No. 3e-11;  
Matches 152; Conservative 124; Mismatches 238; Indels 248; Gaps 28;  
QY 32 PHDPXPPAEQXXLLPXEDCTFYCEYGLKFIAPRDCAPGTEFKFSATCVHAALAGCT 91  
DB 288 PHHDTP-----CTK-----KTTTSTCTKTTTPVTPSSSTTESSAP 328  
QY 92 LFGPAETIQ---APATTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOA 142  
DB 329 VPTPSSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPVTSSTTESSAPVTSST 388  
QY 143 TTTQAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOAQTTOA 196  
DB 389 ESSAPVTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTS 448  
QY 196 -TQAATTPAATTPAATTPAATTP--AATTPAATTPGVAP-----TSAPVWPVICEL 245  
DB 449 STTESSAPVTPSSSTTESSAPVTSSTTESSAPVTSSTTESSAPVTPSSSTTESSAPV- 501  
QY 245 LPNGCPADFIDHLLPHDKYCNLFYQCSNGYTFEQRCPEGLYFNVQRCDSANVECDG 305  
DB 501 -----TSSTTESSAP-----TSSTTESSAP-----VPTPSSSTTES 522  
QY 306 EISAP--PVTEGNEDEDIDGLLDNGCPANFEIDWLLPHGNRCDKYQCVHGNLVERRC 364  
DB 523 SSAPATTPSSSTTESSAPV----- 543  
QY 365 GAGTHFSFELQOCDHIELVGCPLPGGESEEDVDEDACTGWYCPTEPIEWELPNGCPAD 424

EUKARYOTA; FUNGI; ASCOMYCOTINA; HEMIASCOCYCETES.

[1]

SEQUENCE FROM N.A.

STRAIN-S288C / AB972;

BARRELL B.G., BADCOCK K., BANKIER A.T., BOWMAN S., BROWN D.,

CHURCHER C.M., CONNOR R., COPSEY T., DEAR S., DEVLIN K., FRASER A.,

GENTLES S., HAMLYN N., HORSNELL T.S., HUNT S., JAGELS K., JONES M.,

LOUIS E., LYE G., MOULE S., MOULE T., ODELL C., PEARSON D.,

RAJANDREAM M.A., RILES L., ROWLEY N., SKELTON J., SMITH V.,

WALSH S.V., WHITEHEAD S.;

SUBMITTED (DEC-1994) TO EMBL/GENBANK/DBJ DATA BANKS.

[2]

SEQUENCE OF 1-242 AND 762-1331 FROM N.A.

MEDLINE; 87194600.

YAMASHITA I., NAKAMURA M., FUKUI S.;

J. BACTERIOL. 169:2142-2149(1987).

[3]

SEQUENCE OF 1-31 FROM N.A.

STRAIN-SPX101-1C;

MEDLINE; 89031230.

PADO J.M., IANEZ E., ZALACAIN M., CLAROS M.G., JIMENEZ A.;

FEBS LETT. 239:179-184(1988).

-1- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL 1,4-LINKED ALPHA-D-

GLUCOSE RESIDUES SUCCESSIVELY FROM NON-REDUCING ENDS OF THE CHAINS

WITH RELEASE OF BETA-D-GLUCOSE.

EMBL; Z47047; G763364; -

EMBL; Z38061; G557822; -

EMBL; M16164; G172525; -

EMBL; M16165; G172526; -

EMBL; X13857; G4552; -

PIR; B26877; B26877.

PIR; A26877; A26877.

PIR; S48478; S48478.

SGD; L0002104; STAL.

HYDROLASE; GLYCOSIDASE; POLYSACCHARIDE DEGRADATION; GLYCOPROTEIN;

SIGNAL; MULTIGENE FAMILY.

SIGNAL 1 21

CHAIN 22 1367

DOMAIN 210 1367

CARBOHYD 817 817

CARBOHYD 874 874

SEQUENCE 1367 AA; 136110 MW; 212ED195 CRC32;

Query Match

Best Local Similarity 9.3%; Score 410; DB 1; Length 1367;

Matches 152; Conservative 124; Mismatches 238; Indels 248; Gaps 28;

```

Db 543 -----TSSTESSAPV-----PTPSSSTTE 563
QY 425 FSDHLLPHESGQYLOCVHGTIARPCGNLHFSPTQSCSPVTAQGVFCDSDNQ 484
Db 564 SS-----STPVT-----SSTESSAPVPT-----PSSST 588
QY 485 CYSSTAAPTAA-----TAAPTAATAAPTAAAPSVVPPATPPATAAPV-----PT 530
Db 589 TESSAPVPTPSSSTESSAPAPTSSSTESSAPVTSSTESSAPVPTPSSSTTES 648
QY 531 TAIPTAAPTAAAPTAAAPESPTVTVPTAAAPTAAAPTAAAPTAAAPTAAAPT 589
Db 649 SNAVPTPSSSTESSAPVPTPSSSTESSAPVTSSTESSAPVTSSTESSAPV 707
QY 590 AAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 645
Db 708 PTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVTSSTESSAP 763
QY 646 NTVTAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 700
Db 764 VPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESSAPVPTPSSSTESS 823
QY 701 ATP-----EDDDIDPLND-----PINPCVECNVLP 729
Db 824 SSTPSSSTESSAPVPTPSSSTESSAPVTSSTESSAPVPTPSSSTESSAP 865

RESULT 7
SLP1_CLOTH
ID SLP1_CLOTH STANDARD; PRT; 1664 AA.
AC Q06852;
DT 01-JUN-1994 (REL. 29, CREATED)
DT 01-JUN-1994 (REL. 29, LAST SEQUENCE UPDATE)
DT 01-FEB-1996 (REL. 33, LAST ANNOTATION UPDATE)
DE CELL SURFACE GLYCOPROTEIN 1 PRECURSOR (OUTER LAYER PROTEIN B) (S-LAYER
DE PROTEIN, 1).
GN OLPB.
OS CLOSTRIDIUM THERMOCELLUM.
OC PROKARYOTA; FIRMICUTES; ENDOSPORE-FORMING RODS AND COCCI; BACILLACEAE.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIB 10682.
RX MEDLINE: 93209931.
RA FUJINO T., BEGUIN P., AUBERT J.-P.;
J. BACTERIOL. 175:1891-1899(1993).
CC -!- SUBUNIT: ASSEMBLED INTO MONO-LAYERED CRYSTALLINE ARRAYS.
CC -!- SUBCELLULAR LOCATION: CELL WALL.
CC -!- SIMILARITY: CONTAINS 4 S-LAYER HOMOLOGY (SLH) DOMAINS.
DR EMBL; X67506; G296881; -.
DR PROSITE; PS01072; SLH_DOMAIN; 2.
KW CELL WALL; S-LAYER; SIGNAL; REPEAT.
FT SIGNAL 1 28 POTENTIAL.
FT CHAIN 29 1664 CELL SURFACE GLYCOPROTEIN 1.
FT DOMAIN 36 763 4 X 156 AA APPROXIMATE REPEATS.
FT REPEAT 36 191 1.
FT REPEAT 207 363 2.
FT REPEAT 409 565 3.
FT REPEAT 607 763 4.
FT DOMAIN 771 1377 APPROXIMATE TANDEM REPEATS OF
T-P-S-D-E-P.
FT DOMAIN 1378 1449 GLY/PRO/SER/THR-RICH.
FT DOMAIN 1453 1494 SLH 1 (INCOMPLETE).
FT DOMAIN 1495 1565 SLH 2.
FT DOMAIN 1566 1625 SLH 3.
FT DOMAIN 1626 1646 SLH 4 (INCOMPLETE).
FT SEQUENCE 1664 AA; 178194 MW; BFC7D6A CRC32;

```

Query Match 8.5%; Score 372.5; DB 1; Length 1664;  
 Best Local Similarity 23.5%; Pred. No. 1.3e-09;  
 Matches 176; Conservative 66; Mismatches 325; Indels 183; Gaps 30;

```

QY 18 PEVSDAENKPAHPHDPXPAEQXLLPXEYDCFKFYCYCEYGLKFIAPRD-CAPGTETK 76
Db 825 PTPSD-EPTPS-DEPTPSDEPTSPSETP-----EETPTDTPSDEPTPSDE-- 874
QY 77 FSAQICVHAALAGCTLPGPAAETTOAPATTOAPTTOAPTTOAPTTOAPTTOAPT 136
Db 874 -----PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEP 917
QY 137 TTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPT 192
Db 918 TPDSEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 977
QY 193 PTITTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPTTOAPT 251
Db 978 TPDSE--PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1033
QY 252 ADFDHLHLPHKYNLFQCSNGVTFEQRCEGLVFNFYVQRCDSPANVECDGETSPAP 311
Db 1034 TPDSEPTPSD-----EPTPSETPEPIP-----TDTPSD-----EPTPSD 1068
QY 312 PVTEGNEDEDIDGDLNCGPANEIDWLLPHGNCRDKYQCQVHGNLVERRCGAGTHFS 371
Db 1069 EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1106
QY 372 FELQOCDHIELVGCCTLPGESEEVDEDACTGWYCPTPIEWELPNCPADFSIDHLL 431
Db 1106 -----EPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1149
QY 432 PHESDCGOVLCVHGTIARPCGNLHFSPTQSCSPVTAQGVFCDSDNQCTSTAAP 491
Db 1150 PTPSD-----EPTPSD-----EPTPSDEPTP-----SDEPTSETPE 1181
QY 492 TAAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 547
Db 1182 EPIPTDTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDE 1241
QY 548 TAAPESPTTVTV-PTAAPTAAAPTAAVPEIPITVTSAPTAAPTAAAPTAAAPTAVPE 606
Db 1242 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1297
QY 607 IPTTVTSP-----PTAAPTAAAPTAAVTVVPTAA-----PTTA-- 642
Db 1298 TPDSEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1357
QY 642 -APAPNTTVTAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 1357
Db 1358 PTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPTPSDEPT 1417
QY 687 PNTTAAPTVTTTSAPATTPEDDDDIDPPLND 716
Db 1418 PTP-----TPTSKPTSTAPATEIEEPTPSD 1442

RESULT 8
ANP_NOTCO
ID ANP_NOTCO STANDARD; PRT; 822 AA.
AC P24856;
DT 01-MAR-1992 (REL. 21, CREATED)
DT 01-NOV-1995 (REL. 32, LAST SEQUENCE UPDATE)
DT 01-NOV-1997 (REL. 35, LAST ANNOTATION UPDATE)
DE ANTIFREEZE GLYCOPOLYMER POLYPROTEIN AFGP7/AFGP8 PRECURSOR.
GN AFGP8.
OS NOTOTHEMIA CORIICES NEGLECTA (BLACK ROCKCOD) (YELLOWBELLY ROCKCOD).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; PISCES; GNATHOSTOMATA;
OC OSTEICHTHES; ACTINOPTERYGII; PERCIFORMES.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 91067687.
RA HSIAO K.-C., CHENG C.-H.C., FERNANDES I.E., DETRICH H.W. III,
DEVRIES A.L.;
PROC. NATL. ACAD. SCI. U.S.A. 87:9265-9269(1990).
CC -!- FUNCTION: ANTIFREEZE PROTEINS LOWER THE BLOOD FREEZING POINT.
CC -!- TISSUE SPECIFICITY: SYNTHESIZED BY THE LIVER AND SECRETED INTO

```

CC THE BLOOD FROM WHICH THEY BECOME DISTRIBUTED TO ALMOST THE ENTIRE  
CC EXTRACELLULAR SPACE.

CC -1- PTM: THE DISACCHARIDE GALACTOSE-N-ACETYLGLALACTOSAMINE IS ATTACHED  
CC TO THREONINES IN AFGP8 AND AFGP7.

CC -1- CONTAINS 44 COPIES OF AFGP8 AND TWO COPIES OF AFGP7.

DR EMBL: M55000; G213392; -

DR PIR: A38420; A38420.

DR HSP: P04002; IATF.

KW ANTIFREEZE PROTEIN; GLYCOPROTEIN; POLYPROTEIN; REPEAT;

FT SIGNAL 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 92 93 94 95 96 97 98 99 100 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 119 120 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 139 140 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 159 160 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 179 180 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 199 200 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 219 220 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 239 240 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258 259 260 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278 279 280 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298 299 300 301 302 303 304 305 306 307 308 309 310 311 312 313 314 315 316 317 318 319 320 321 322 323 324 325 326 327 328 329 330 331 332 333 334 335 336 337 338 339 340 341 342 343 344 345 346 347 348 349 350 351 352 353 354 355 356 357 358 359 360 361 362 363 364 365 366 367 368 369 370 371 372 373 374 375 376 377 378 379 380 381 382 383 384 385 386 387 388 389 390 391 392 393 394 395 396 397 398 399 400 401 402 403 404 405 406 407 408 409 410 411 412 413 414 415 416 417 418 419 420 421 422 423 424 425 426 427 428 429 430 431 432 433 434 435 436 437 438 439 440 441 442 443 444 445 446 447 448 449 450 451 452 453 454 455 456 457 458 459 460 461 462 463 464 465 466 467 468 469 470 471 472 473 474 475 476 477 478 479 480 481 482 483 484 485 486 487 488 489 490 491 492 493 494 495 496 497 498 499 500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524 525 526 527 528 529 530 531 532 533 534 535 536 537 538 539 540 541 542 543 544 545 546 547 548 549 550 551 552 553 554 555 556 557 558 559 560 561 562 563 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596 597 598 599 600 601 602 603 604 605 606 607 608 609 610 611 612 613 614 615 616 617 618 619 620 621 622 623 624 625 626 627 628 629 630 631 632 633 634 635 636 637 638 639 640 641 642 643 644 645 646 647 648 649 650 651 652 653 654 655 656 657 658 659 660 661 662 663 664 665 666 667 668 669 670 671 672 673 674 675 676 677 678 679 680 681 682 683 684 685 686 687 688 689 690 691 692 693 694 695 696 697 698 699 700 701 702 703 704 705 706 707 708 709 710 711 712 713 714 715 716 717 718 719 720 721 722 723 724 725 726 727 728 729 730 731 732 733 734 735 736 737 738 739 740 741 742 743 744 745 746 747 748 749 750 751 752 753 754 755 756 757 758 759 760 761 762 763 764 765 766 767 768 769 770 771 772 773 774 775 776 777 778 779 780 781 782 783 784 785 786 787 788 789 790 791 792 793 794 795 796 797 798 799 800 801 802 803 804 805 806 807 808 809 810 811 812 813 814 815 816 817 818 819 820 821 822 823 824 825 826 827 828 829 830 831 832 833 834 835 836 837 838 839 840 841 842 843 844 845 846 847 848 849 850 851 852 853 854 855 856 857 858 859 860 861 862 863 864 865 866 867 868 869 870 871 872 873 874 875 876 877 878 879 880 881 882 883 884 885 886 887 888 889 890 891 892 893 894 895 896 897 898 899 900 901 902 903 904 905 906 907 908 909 910 911 912 913 914 915 916 917 918 919 920 921 922 923 924 925 926 927 928 929 930 931 932 933 934 935 936 937 938 939 940 941 942 943 944 945 946 947 948 949 950 951 952 953 954 955 956 957 958 959 960 961 962 963 964 965 966 967 968 969 970 971 972 973 974 975 976 977 978 979 980 981 982 983 984 985 986 987 988 989 990 991 992 993 994 995 996 997 998 999 1000

QY 187 PTTQAPTITQIA-----ATTPA-ATTP-----AATTPAATTPAATTPA---ATTPGVA 232  
Db 141 ATAATAATPATPAFHAATAATATAATPALIFAATAATAATPA--TPAFHAATAATPA 198  
QY 233 PPSAPVWPICELLPNGCPA-----DFDIHLLIPHDYKYNLFYOCSSNGYFQRCPLGLYF 288  
Db 199 TAATPALIFAATAATAATATAATPALIFAATAATAATATAATPALIFAATAATAATATAALNF 258  
QY 289 NPVQRCDSANVECDGEISPAAPPTEGNEDEDEDIDGLDNGCPANFEIDWLLPHGNCRC 348  
Db 259 AATAATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATAATATA 290  
QY 349 DKYIQCVHGNLVERRGAGTHFSFELQOCDHIELVGTCLPGGESEEDVDEDACTGWYCP 408  
Db 290 -----LI-----FAA 294  
QY 409 TEPIEWEPLNGCPADFSIDHLLPHESDCGQYLOQVGHGTIARCPGNLHF-----SPAT 463  
Db 295 TAATAATPATPAAC--NFAATAATPATPAATATAATATAATATAATATAATATAATATAATPAT 352  
QY 464 -----QSCSPVTAGCVFECSDNOCTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 511  
Db 353 AATPALIFAATAATAATATAAL-----NFAATAATATAATATAATATAATATAATATAAT 404  
QY 512 -----STVPPATPP-----ATAAPVPTTATPTP--APTAAPTAAPTAAPTAAPESPT 555  
Db 405 ATAALNFAATAATAATPATPAFNFATAATATAATATAATATAATATAATATAATATAALNFAA 464  
QY 556 TTVTPPTAAPT-----APTAVPEIPTVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAP 612  
Db 465 TAATPATATAATPALIFAATAA-----TAATAATAALHFAATAATAATATAALNFAATAA 518  
QY 613 SPPTA-----APTAAAPAP-----NTTVTPPTAAPTATAAPNTTVTAPPTAA-PT 658  
Db 519 TPATAATPALIFAATAATAATATAAFNFAATAATAATATAALNFAATAATAATATAATATA 578  
QY 659 TAAPAPNTTVTPPTA-----APTAAPPTVAHAPNTTAAAPVTTTSAPATTP 704  
Db 579 LIFAATAATAATATAALNFAATAATATAATATAATATAATATAATATAATATAATATAAT 628  
RESULT 9  
VTP3-TTV1V STANDARD; PRT; 474 AA.  
AC P19275;  
DT 01-NOV-1990 (REL. 16, CREATED)  
DT 01-NOV-1990 (REL. 16, LAST SEQUENCE UPDATE)  
DT 01-FEB-1991 (REL. 17, LAST ANNOTATION UPDATE)  
DE VIRAL PROTEIN TPX.  
OS THERMOPROTEUS TENAX VIRUS 1 (STRAIN VT3) (TTV1).  
OC VIRIDAE; DS-DNA ENVELOPED VIRUSES; LIPOTHRIXVIRIDAE.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 90245666.  
RA NEUMANN H., ZILLIG W.;  
RL NUCLEIC ACIDS RES. 18:2171-2171(1990).  
DR EMBL; X14717; E1192689; -  
DR PIR; S06686; S06686.  
KW REPEAT.  
KW REPEAT.  
FT DOMAIN 278 467 3 THR-PRO REPEATS REGIONS AND TWO NEAR  
FT REPEAT 278 367 IDENTICAL REPEATS.  
FT REPEAT 368 377 THR-PRO(N).  
FT REPEAT 378 436 THR-PRO(N).  
FT REPEAT 437 446 THR-PRO(N).  
FT REPEAT 447 467 THR-PRO(N).  
SQ SEQUENCE 474 AA; 49507 MW; BAC5D831 CRC32;

Query Match 8.4%; Score 367.5; DB 1; Length 474;  
Best Local Similarity 32.6%; Pred. No. 8.5e-10;

Query Match 8.4%; Score 369.5; DB 1; Length 822;  
Best Local Similarity 25.9%; Pred. No. 1e-09;  
Matches 184; Conservative 23; Mismatches 314; Indels 189; Gaps 25;

QY 82 CVHAALAGC-----TLGPPPAETTQAPATTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 136

Db 21 CVCVSVCVCTAVTAAPAAATAATAATATAALNFAATAATATAATATAATATAATATAAT 80

QY 137 TTTOA---PTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 186

Db 81 PATAALNFAATAATATAATPALIFAFAAATAATATAALNFAATAATATAATATAATPALIFA 140

[illegible]



[illegible]



GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 02:42:01 ; Search time 32.5 seconds  
(without alignments)  
1369.893 Million cell updates/sec

Title: US-09-294-663-4

Perfect score: 4395

Sequence: 1 MIKTLFLTLGLVAARPEV.....VFIPWNKLDIROLNLFEL 807

Scoring table: BLOSUM62

Searched: 180763 seqs, 55169189 residues

Database :

- SPTREMBL\_8.\*
- 1: sp\_fungi.\*
  - 2: sp\_human.\*
  - 3: sp\_invertebrate.\*
  - 4: sp\_mammal.\*
  - 5: sp\_mhc.\*
  - 6: sp\_organelle.\*
  - 7: sp\_phase.\*
  - 8: sp\_plant.\*
  - 9: sp\_bacteria.\*
  - 10: sp\_rodent.\*
  - 11: sp\_virus.\*
  - 12: sp\_vertebrate.\*
  - 13: sp\_unclassified.\*
  - 14: sp\_archaea.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4375	99.5	807	3 O18511	O18511 trichoplusi
2	4266.5	97.1	788	3 O18510	O18510 trichoplusi
3	542	12.3	1404	2 Q92954	Q92954 homo sapien
4	481	10.9	851	3 O17893	O17893 caenorhabdi
5	477.5	10.9	867	11 Q39782	Q39782 equine herp
6	473.5	10.8	3570	2 Q99552	Q99552 homo sapien
7	472.5	10.8	866	11 Q39781	Q39781 equine herp
8	459	10.4	801	3 Q23635	Q23635 caenorhabdi
9	453	10.3	770	3 Q23098	Q23098 caenorhabdi
10	431	9.8	2378	3 P91365	P91365 caenorhabdi
11	422.5	9.6	373	3 Q76810	Q76810 anopheles g
12	421.5	9.6	1188	8 Q41805	Q41805 zea mays (m
13	414.5	9.4	825	3 Q17921	Q17921 caenorhabdi
14	412.5	9.4	1611	1 O42854	O42854 schizosacch
15	405	9.2	507	12 O13028	O13028 boreogadus
16	403	9.2	622	2 Q14881	Q14881 homo sapien
17	398	9.1	750	11 Q39307	Q39307 equine herp
18	396.5	9.0	1777	3 Q22579	Q22579 caenorhabdi
19	390.5	8.9	1151	12 Q57580	Q57580 gallus gall
20	384	8.7	2187	10 P70670	P70670 mus musculu
21	380.5	8.7	846	3 O01699	O01699 caenorhabdi
22	373	8.5	1118	4 Q28226	Q28226 canis famil
23	370.5	8.4	412	3 Q77242	Q77242 heterodera
24	368	8.4	1131	3 Q18529	Q18529 caenorhabdi
25	365	8.3	353	3 Q25677	Q25677 podocoryne
26	348.5	7.9	279	2 Q14888	Q14888 homo sapien
27	347.5	7.9	350	8 Q41719	Q41719 zea diplope
28	346.5	7.9	5376	10 Q88799	Q88799 mus musculu
29	346	7.9	722	12 O13083	O13083 dissostichu

## ALIGNMENTS

RESULT 1

O18511  
ID O18511 PRELIMINARY; PRT: 807 AA.  
AC O18511;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE INSECT INTESTINAL MUCIN IIM22.  
OS TRICHOPLUSIA NI (CABBAGE LOOPER).  
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
OC PTERYGOTA; LEPIDOPTERA; NOCTUIDAE; NOCTUIDAE; TRICHOPLUSIA.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97341213.  
RA WANG P., GRANADOS R.R.;  
RT "Molecular cloning and sequencing of a novel invertebrate intestinal  
RT mucin cDNA."  
RL J. BIOL. CHEM. 272:16663-16669(1997).  
DR EMBL: AF000606; G2224921;  
SQ SEQUENCE 807 AA; 83408 MW; 15E08C34 CRC32;

Query Match 99.5%; Score 4375; DB 3; Length 807;  
Best Local Similarity 99.4%; Pred. No. 1.3e-119;  
Matches 802; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Oy 1 MIKTLFLTLGLVAARPEVSDAEKNPALHEPHDPXPPAEQXXLLPXEYDCTKFFYCEYG 60  
Db 1 MIKTLFLTLGLVAARPEVSDAEKNPALHEPHDPXPPAEQXXLLPXEYDCTKFFYCEYG 60  
Oy 61 LKFTAPRDCAPGTEKFSQAOTCVHAALAGCTLPGPAAETTOAPTTOAPTTOAPTTO 120  
Db 61 LKFTAPRDCAPGTEKFSQAOTCVHAALAGCTLPGPAAETTOAPTTOAPTTOAPTTO 120  
Oy 121 APTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQA 180  
Db 121 APTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQA 180  
Oy 181 PTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTT 240  
Db 181 PTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQAPTTTTQA 240  
Oy 241 PICELLPNGCDAFDIHLIIPHDKYCNLFYQCSNGYFQRCPEGLYNPNYVQRCDSAN 300  
Db 241 PICELLPNGCDAFDIHLIIPHDKYCNLFYQCSNGYFQRCPEGLYNPNYVQRCDSAN 300  
Oy 301 VECDEISPPVPTGNEDEDEDIGDLNDCPANFEDWLLPHGNRCDKYQCQVHGNLV 360  
Db 301 VECDEISPPVPTGNEDEDEDIGDLNDCPANFEDWLLPHGNRCDKYQCQVHGNLV 360  
Oy 361 ERRCAGTHFSFELQOCCDHIELVGCTLPFGSEEVVDVEDACTGWYCTEPIEWELPNG 420  
Db 361 ERRCAGTHFSFELQOCCDHIELVGCTLPFGSEEVVDVEDACTGWYCTEPIEWELPNG 420





RL SUBMITTED (SEP-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL: U70136; G1572721; -  
DR PROSITE: PS00524; SOMATOMEDIN\_B; 2.  
DR PFAM: PF00045; hemopexin; 2.  
DR PFAM: PF01033; Somatomedin\_B; 2.  
SQ SEQUENCE 1404 AA; 151090 MW; 6C1F4DEA CRC32;

Query Match 12.3%; Score 542; DB 2; Length 1404;  
Best Local Similarity 30.4%; Pred. No. 1.6e-09;  
Matches 215; Conservative 23; Mismatches 194; Indels 276; Gaps 37;

QY 93 PGPAE-TTQAPATT---QAPITTOQAPITTT---QAPITTOQAPITTTQ-APTITQ-AP 142  
DB 331 PTPKAETTTGKALTTPKEPTTPKEPASTTPKEPTTPKESAPTTTPKEPAPTTKSP 390  
QY 143 TTQ--APTITQAPITTT---QAPTITTO-APTITTOQAPITTTQ-APTITQ-APTITQ 192  
DB 391 TTPKEAPITTTKEPAPTTPKEPAPTTPKESAPTTTPKESAPTTTPKAPPTTPKE 450  
QY 193 PTTITQAAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPA 252  
DB 451 P-----APTTPKEPTP--TTPKEPAPTTPKEPAPTTPKEPAPT-APKKA----- 492  
QY 253 DFDIHLPHDKKYNLFYQCSNGYTEQRCPEGLYFNYPVQRCDSANVECDGEISFAPP 312  
DB 492 -----PTTPKEPAPTTPK-----PTTPKEPAPTTPK-----EPAPT 509  
QY 313 VTGNEDEDIDGLDNGCPANFEIDWLLPHGNCRDKYQCVHGNLVERRCGAGTHFSF 372  
DB 510 TKKE-----PS----- 516  
QY 373 ELQCCDHIELVGLTLPGESEEDVDEDACTGWYCPTIEWEPLP---NGCPADFSIDH 429  
DB 516 -----PTTP--KEPAPTTPKSAFT----- 533  
QY 430 LLPHSDCGOYLOCVHGOIARPCPNLHFSPATQCSFSPVAGCQVFECDSNQCTSTA 489  
DB 533 -----TTKEPAPTTPKSAFTTPKEPSPTT-----TKEP 560  
QY 490 APTA-----APTAAAPTAAAPTAAPTSTVVPATPATAPATAPATAPATAPATAPAT 545  
DB 561 APTTPKEPAPTTPKAPPTTPKEPAPTTPKEPA-PTTTKKAP--TAPKEPAPTTPKEA 617  
QY 546 PTTAAPESEPT-----VTVPPTAAPT-----APTAVPEIPIVT---SAPTAAPTAA 593  
DB 618 PTPPKLITTPPEKLAPTTPKEPAPTTPPELAPT--PEEPTTTPKEPAPTTPKAAAPN 675  
QY 594 APTAAPTAVPEIPT-----VTSPTAAPT-----AAPAPT----- 629  
DB 676 TPKEPAPTTPKEPAPTTPKEPAPTTPKEAPTTPKGTAPTTPKEPAPTTPKAPKELAP 735  
QY 629 TVTVVPPTAAPTAAAP-----NTTVTAPTAAPT--AAPAPTNTTVPTAAPTAPPTV 683  
DB 736 TTTKEPTST-TSDKPAPTTPKGTAPTTPKEPAPTTPKEPAPTTPKGTAPTTPKEPAPTTP 794  
QY 684 AH-----APNTAAPTAVTTTS-----APATTPEDDDIDP--PLNDPIMP 720  
DB 795 KKPAPKELAPTTTKGTSITSDKPAPTTPKAPTTPKAPTTPKAPTTPKPP 842

RESULT 4  
O17893 PRELIMINARY; PRT; 851 AA.  
ID O17893;  
AC O17893;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)  
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE F55811.3.  
OS CAENORHABDITIS ELEGANS.  
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTERA; RHABDITIDA; RHABDITIDAE;  
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.  
RN [1]

RP SEQUENCE FROM N.A.  
RX MEDLINE; 94150718.  
RA WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
RA BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,  
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SMALDON N., SMITH A.,  
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
RA WILKINSON-SPROAT J., WOHLDMAN P.,  
RT \*2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
RT elegans.\*;  
RL NATURE 368:32-38(1994).  
DR EMBL: Z83318; E1186497; -  
SQ SEQUENCE 851 AA; 92127 MW; 87CF937F CRC32;

Query Match 10.9%; Score 481; DB 3; Length 851;  
Best Local Similarity 25.0%; Pred. No. 6.3e-08;  
Matches 160; Conservative 22; Mismatches 150; Indels 308; Gaps 16;  
QY 91 TLPGPAAETTOQAPATTQAPITTTQAPITTTQAPITTTQAPITTTQAPITTTQAPITTT 150  
DB 320 TRPTIPTKLITTPSEL-STTTASVPTTTTSVPTTTTTPVTPTT---TTTVSTTT 374  
QY 151 TQAPTITTTQAPITTTQAPITTTQAPITTTQAPITTTQAPITTTQAPITTTQAPITTT 210  
DB 375 TVPPTTTTTPVTPTTTTSTTTTTPVT---TTVPTTTTTPVTPTT---TVPTT 422  
QY 211 TTPAATTPAATTPAATTPGVPAPTAPVWPPICELLPNCPADFDIHLPHDKKYNLFY 270  
DB 423 TTSVPTT--TTTVPTTTTTPVTPTT----- 446  
QY 271 QCSNGYTFEQRCEGLYFNYPVQRCDSANVECDGEISFAPPVTEGNEDEDIDGLDLDN 330  
DB 446 ----- 446  
QY 331 GCPANFEIDWLLPHGNCRDKYQCVHGNLVERRCGAGTHFSFELQCCDHIELVGLTLP 390  
DB 446 ----- 446  
QY 391 ESEEDVDEDACTGWYCPTIEWEPLPNCGPADFSIDLPHSPHSDCGOYLOCVHGOI 450  
DB 446 -----TVS 448  
QY 451 RCPGNLHFSPATQCSFSPVAGCQVFECDSNQCTSTAAPTAAAPTAAAPTAAAPTAA 510  
DB 449 -----TTTTPVTPTTTTTPVTPTT---TTTV 468  
QY 511 PSTVVPPTAAPTAAAPVPTTALPTAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPT 570  
DB 469 PTTT---TTVPTTTTTPVTPTT---TTVPTT--TSVPTTTTTPVTPTTTTTPVTPTT 521  
QY 571 TAVPEIPIVTISAPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 630  
DB 522 TTV---PTTTTTPVTPTTTTTPVTPTTTTTPVTPTTTTTPVTPTTTTTPVTPTT 578  
QY 631 TVPPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 690  
DB 579 TVPTT---TTTVPTTTTTPVTPTTTTTPVTPTTTTTPVTPTTTTTPVTPTT---TT 620  
QY 691 AAPVTIT-----SAPAITPEDDDIDPPLNDPIMPVCEE 724  
DB 621 TVPTTTTTPVLFVYTEPTTDYDVLLEASMEN---STCGEE 657

RESULT 5  
O39782 PRELIMINARY; PRT; 867 AA.  
ID O39782;  
AC O39782;  
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)



Db	QY	Seq1	Seq2	Seq3	Seq4	Seq5	Seq6	Seq7	Seq8	Seq9	Seq10	Seq11	Seq12	Seq13	Seq14	Seq15	Seq16	Seq17	Seq18	Seq19	Seq20	Seq21	Seq22	Seq23	Seq24	Seq25	Seq26	Seq27	Seq28	Seq29	Seq30	Seq31	Seq32	Seq33	Seq34	Seq35	Seq36	Seq37	Seq38	Seq39	Seq40	Seq41	Seq42	Seq43	Seq44	Seq45	Seq46	Seq47	Seq48	Seq49	Seq50	Seq51	Seq52	Seq53	Seq54	Seq55	Seq56	Seq57	Seq58	Seq59	Seq60	Seq61	Seq62	Seq63	Seq64	Seq65	Seq66	Seq67	Seq68	Seq69	Seq70	Seq71	Seq72	Seq73	Seq74	Seq75	Seq76	Seq77	Seq78	Seq79	Seq80	Seq81	Seq82	Seq83	Seq84	Seq85	Seq86	Seq87	Seq88	Seq89	Seq90	Seq91	Seq92	Seq93	Seq94	Seq95	Seq96	Seq97	Seq98	Seq99	Seq100	Seq101	Seq102	Seq103	Seq104	Seq105	Seq106	Seq107	Seq108	Seq109	Seq110	Seq111	Seq112	Seq113	Seq114	Seq115	Seq116	Seq117	Seq118	Seq119	Seq120	Seq121	Seq122	Seq123	Seq124	Seq125	Seq126	Seq127	Seq128	Seq129	Seq130	Seq131	Seq132	Seq133	Seq134	Seq135	Seq136	Seq137	Seq138	Seq139	Seq140	Seq141	Seq142	Seq143	Seq144	Seq145	Seq146	Seq147	Seq148	Seq149	Seq150	Seq151	Seq152	Seq153	Seq154	Seq155	Seq156	Seq157	Seq158	Seq159	Seq160	Seq161	Seq162	Seq163	Seq164	Seq165	Seq166	Seq167	Seq168	Seq169	Seq170	Seq171	Seq172	Seq173	Seq174	Seq175	Seq176	Seq177	Seq178	Seq179	Seq180	Seq181	Seq182	Seq183	Seq184	Seq185	Seq186	Seq187	Seq188	Seq189	Seq190	Seq191	Seq192	Seq193	Seq194	Seq195	Seq196	Seq197	Seq198	Seq199	Seq200	Seq201	Seq202	Seq203	Seq204	Seq205	Seq206	Seq207	Seq208	Seq209	Seq210	Seq211	Seq212	Seq213	Seq214	Seq215	Seq216	Seq217	Seq218	Seq219	Seq220	Seq221	Seq222	Seq223	Seq224	Seq225	Seq226	Seq227	Seq228	Seq229	Seq230	Seq231	Seq232	Seq233	Seq234	Seq235	Seq236	Seq237	Seq238	Seq239	Seq240	Seq241	Seq242	Seq243	Seq244	Seq245	Seq246	Seq247	Seq248	Seq249	Seq250	Seq251	Seq252	Seq253	Seq254	Seq255	Seq256	Seq257	Seq258	Seq259	Seq260	Seq261	Seq262	Seq263	Seq264	Seq265	Seq266	Seq267	Seq268	Seq269	Seq270	Seq271	Seq272	Seq273	Seq274	Seq275	Seq276	Seq277	Seq278	Seq279	Seq280	Seq281	Seq282	Seq283	Seq284	Seq285	Seq286	Seq287	Seq288	Seq289	Seq290	Seq291	Seq292	Seq293	Seq294	Seq295	Seq296	Seq297	Seq298	Seq299	Seq300	Seq301	Seq302	Seq303	Seq304	Seq305	Seq306	Seq307	Seq308	Seq309	Seq310	Seq311	Seq312	Seq313	Seq314	Seq315	Seq316	Seq317	Seq318	Seq319	Seq320	Seq321	Seq322	Seq323	Seq324	Seq325	Seq326	Seq327	Seq328	Seq329	Seq330	Seq331	Seq332	Seq333	Seq334	Seq335	Seq336	Seq337	Seq338	Seq339	Seq340	Seq341	Seq342	Seq343	Seq344	Seq345	Seq346	Seq347	Seq348	Seq349	Seq350	Seq351	Seq352	Seq353	Seq354	Seq355	Seq356	Seq357	Seq358	Seq359	Seq360	Seq361	Seq362	Seq363	Seq364	Seq365	Seq366	Seq367	Seq368	Seq369	Seq370	Seq371	Seq372	Seq373	Seq374	Seq375	Seq376	Seq377	Seq378	Seq379	Seq380	Seq381	Seq382	Seq383	Seq384	Seq385	Seq386	Seq387	Seq388	Seq389	Seq390	Seq391	Seq392	Seq393	Seq394	Seq395	Seq396	Seq397	Seq398	Seq399	Seq400	Seq401	Seq402	Seq403	Seq404	Seq405	Seq406	Seq407	Seq408	Seq409	Seq410	Seq411	Seq412	Seq413	Seq414	Seq415	Seq416	Seq417	Seq418	Seq419	Seq420	Seq421	Seq422	Seq423	Seq424	Seq425	Seq426	Seq427	Seq428	Seq429	Seq430	Seq431	Seq432	Seq433	Seq434	Seq435	Seq436	Seq437	Seq438	Seq439	Seq440	Seq441	Seq442	Seq443	Seq444	Seq445	Seq446	Seq447	Seq448	Seq449	Seq450	Seq451	Seq452	Seq453	Seq454	Seq455	Seq456	Seq457	Seq458	Seq459	Seq460	Seq461	Seq462	Seq463	Seq464	Seq
----	----	------	------	------	------	------	------	------	------	------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	--------	-----



RN SEQUENCE FROM N.A.  
 RP STRAIN-BRISTOL N2;  
 RX MEDLINE; 94150718.  
 RA WILSON R., AINSOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,  
 RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,  
 RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,  
 RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,  
 RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,  
 RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWKNEEN R., SWALDON N., SMITH A.,  
 RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,  
 RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,  
 RA WILKINSON-SPROAT J., WOHLDMAN P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.  
 RT elegans.";   
 RL NATURE 368:32-38(1994).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA GEISEL C., GATTUNG S.;  
 RL SUBMITTED (JAN-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (DEC-1996) TO EMBL/GENBANK/DBJ DATA BANKS.  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-BRISTOL N2;  
 RA WATERSTON R.;  
 RL SUBMITTED (APR-1997) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; U08046; G1707247;  
 SQ SEQUENCE 2378 AA; 225242 MW; 765EDB23 CRC32;

Query Match 9.8%; Score 431; DB 3; Length 2378;  
 Best Local Similarity 25.1%; Pred. No. 3.4e-06;  
 Matches 158; Conservative 50; Mismatches 144; Indels 278; Gaps 26;  
 QY 111 TTAQPTTTTQAAPTPTTQAAPTPTTQ-----APTPTTQAAPTPTTQAAPTPTTQAAPTPTT 165  
 DB 223 TILGTTTTITTTTIVVASTQTQAPASPTTQAPAT--SPTTQAPATS-----PTTQ 276  
 QY 166 QAPTTTTQAAPTPTTQAAPTPTTQAAPTPTTQAAPTPTTQAAPTPTTQAAPTPTTQAAPTPTT 225  
 DB 277 PAPAT--SPTTQAPAT--SPTTQAPAT--TSPTTQAPATSPTTQAPATSPTTQAPAP 329  
 QY 225 ATTPGV--PAPTSAPVWPPICELLPGCPADFDIHLIPHDKVCNLFYQCSNGYTFEQRCP 283  
 DB 330 ATSPTTQAPATSPTTQ----- 348  
 QY 284 EGLYENPYVQRCDSPANVECDGEISPAPVTEGNEDEDIDGLDNGCPANFEIDWLLP 343  
 DB 348 -----TFA-----TSPTT----- 356  
 QY 344 HGNRCDKYQCVHGNLVERRCGAGTHFSFELQCDHIELVGLTLPGESEEDVDVDEACT 403  
 DB 356 ----- 356  
 QY 404 GWYCETPEIEWEPLNGCPADFSIDHLLPHESDCGQYLOCVHGQIARPCPNLHFSPAT 463  
 DB 356 -----QPTPATSP-----TOPTPAT-----SPTT 375  
 QY 464 QSCSPVTAGCQVFECDSDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 523  
 DB 376 Q--PAPATSP-----ITQAPATSPTTQ--APATSPTTQAPATSPTTQAPAP 419  
 QY 523 -----TAAVPPTTAIPTAPTAAPTAAPTAAPTPTTVPPTTAAPTAAPTAAPTAVPEIPI 578  
 DB 420 ATSPTTQAPATSPTTQAPATSPTTQPTPGI--SPTTQAPAT--SPTTQPSFA----- 471  
 QY 579 TVTSAPTAAPTAAPTAAPTAAPTAVPEIPTTTSPTTAAPTAAPTAAPTNTTVPPTTAA 638

DB 471 ---TNPPTQPSVSSASTQAPATS-----GTTQAPATSATQAPATSATQAPATSATQAPATS 523  
 QY 639 TTAAPAP-----NTTVPAPTAAPTTAAAPA---PNTTVPPTTAAAPTAAAPT 681  
 DB 524 TT-QPAPATSATTQSPVVVTTTGTGTSPTTAGISSSPAATSPSTAVTSPSLGTSSSP- 582  
 QY 682 TVAHAPNTAAPTPTTTS-----APATTP 704  
 DB 582 -LPSSISTSALPIASSASSASSSSPSAASSTTP 610  
 RESULT 11  
 OT6810  
 ID 076810 PRELIMINARY; PRT; 373 AA.  
 AC 076810;  
 DT 01-NOV-1998 (TREMREL. 08, CREATED)  
 DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)  
 DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)  
 DE ICHIT PROTEIN.  
 GN ICHIT.  
 OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).  
 OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;  
 OC PTERYGOTA; DIPTERA; NEMATOCERA; CULICOIDEA; CULICIDAE; ANOPHELES.  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-SUAKOKO; TISSUE-MAINLY MIDGUT;  
 RA DIMOPOULOS G., SEELEY D., WOLF A., KAFATOS F.C.;  
 RT "Malaria infection of the mosquito Anopheles gambiae activates immune  
 RT responsive genes during critical transition stages of the parasite  
 RT life cycle.";   
 RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
 DR EMBL; AJ010903; E1320052;  
 SQ SEQUENCE 373 AA; 40647 MW; A7231CDB CRC32;

Query Match 9.6%; Score 422.5; DB 3; Length 373;  
 Best Local Similarity 35.6%; Pred. No. 1.7e-06;  
 Matches 135; Conservative 23; Mismatches 126; Indels 95; Gaps 20;  
 QY 1 MKTLLELTALGL-----VAARPEVSDAEKNPALHEPHDPXPAEQXLLPXRY-----DCT 52  
 DB 1 MIAAMKVVALGLVLLAVSARAEPGEVTPN-----HNCPP--EMQGLPHFYHPTNCS 52  
 QY 53 KFYIC-----EYGLKFIAPRDCAPGTEKFSQAQTCVHAALAGC--TLPGPPAETTOAPA 104  
 DB 53 RFEYCHMRDAWEY-----ECPAGLHFNVAIDVCDPVPNAKCESQSPGDQTTTLRPT 104  
 QY 105 TTQ-APTTOQA-----PTTTTQAAPTPTTQAAPTPTTQAAPTPTTQAAPTPTTQAAPTPTT 150  
 DB 105 TTTLRPTTTTDDWTTTTTTEATTTTKEPTTTTTSAPTTPSQWTDPTITTTTPVWTDPT 164  
 QY 151 TOAPTPTT-----QAPTPTT-----QAPTPTTQAAPTPTTQAAPTPTT-----PTTT 185  
 DB 165 WSAPTITTTWSQPPPTTTTTPVWTDPTATTTTHAPTITTTTWSDLPPPPPTTTTTPVWID 224  
 QY 185 -QAPTITTOAPTITTOAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTPAATTP 233  
 DB 225 PTATTTTHAPTITTTTWSDLPPPPPTTTTTPVWTDPTITTTTDDYTTAYPTTSEPSTPH 284  
 QY 234 TSAPVWPICELLNGCPADFDIHLIPHDKVCNLFYQCSNGYTFEQRCPGLYFNYPVQ 293  
 DB 285 TD-PHCPTTGATLPN-----YWAHGDCSRYGCLGCVKEKPCDGLYWNDOOK 333  
 QY 294 RCDSPANVECD-GEISPAP 311  
 DB 334 RCDSSSSQCGCPDIPAP 352  
 RESULT 12  
 Q41805  
 ID Q41805 PRELIMINARY; PRT; 1188 AA.  
 AC Q41805;





Db 604 ----- 604  
QY 450 ARPCGNLHFSPATOSCESPVTAGCOVFECDSNOCSTAAAPTAAAPTAAAPTAA 509  
Db 604 -----SSTOTAPPTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 626  
QY 510 APSTVVPATPATAAPVPTTIAIPTAPTAAPTAAAPTAAAPTAAAPTAAAPTAA 568  
Db 627 LUTITLTP-TVPVTTVVPSSATVPTTPTTPTTPTTPTTPTTPTTPTTPTTPTT 680  
QY 569 PTAVPEIPITVTSAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 626  
Db 681 PKPLTSPSTVGTSTPTA-----PANLTTPTTA-PVNPTSTTAPTAPVNP 728  
QY 627 NTTVTVP--TAAPTATAAPNTVTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 684  
Db 729 TTAPTVPPVTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 785  
QY 685 HAPNTTAAPVTTTSPATPTPEDDDIDPPLP 714  
Db 786 SOPATQATPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPTTPT 815

## RESULT 14

O42854 PRELIMINARY: PRT: 1611 AA.

AC O42854;  
DT 01-JUN-1998 (TREMBLREL. 06, CREATED)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST SEQUENCE UPDATE)  
DT 01-JUN-1998 (TREMBLREL. 06, LAST ANNOTATION UPDATE)  
DE HYPOTHETICAL 170.5 KD PROTEIN.  
GN SPAC23A1.16.  
OS SCHIZOSACCHAROMYCES POMBE (FISSION YEAST).  
OC EUKARYOTA; FUNGI; ASCOMYCOTA; ARCHIASCOMYCETES;  
OC SCHIZOSACCHAROMYCETALES; SCHIZOSACCHAROMYCETACEAE;  
OC SCHIZOSACCHAROMYCES.  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA MURPHY L., HARRIS D.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-972H-;  
RA WOOD V., BARRELL B.G., RAJANDREAM M.A.;  
RL SUBMITTED (FEB-1998) TO EMBL/GENBANK/DBJ DATA BANKS.  
DR EMBL; AL021813; E1250590;  
KW HYPOTHETICAL PROTEIN.  
SQ SEQUENCE 1611 AA; 170526 MW; 98811CA3 CRC32;

## Query Match

Best Local Similarity 9.4%; Score 412.5; DB 1; Length 1611;  
Matches 186; Conservative 80; Mismatches 324; Indels 243; Gaps 37;

19 EVSDAEKNPALHE----PHDPXPAEQXLLPXEYDCTKFYCEYGLKFIAPRDCAPGTF 75  
492 EOSNAQAQAPSPKEERLPSEPSNQ----PAEYRDTP-----DTPRNIMP----- 534  
76 KFSATCVHAALAGCTLGPAPATTQAPATTQAPATTQAPATTQAPATTQAPATTQ 135  
534 -----LPGLMSADQIKVTPSPNDADKAIVAEGPNNEE-----TKGPVIPETQ 577  
136 PTTTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQAPTPTTQ 192  
538 EISEQOVHKTPSEKQKVLSPPIITNDFKELASNEAHEAVQKPSAPQVTRLMAPQDS 637  
QY 193 PTTITQAATT---PAATTAAATTAATTAATTAATTAATTAATTAATTAATTAAT 249  
Db 638 SSVVTPSPSLDTPARAVRKVID--GIDPPKEAGAGATADVESANSPITPTWHSDF 695  
QY 250 CPAEDF-THLLTP-----HDKYCNLFYQCSNGYTFEQRCRCPGLYFN----- 290

Db 696 TSKSEPIERKLPSRISVETEDSIDEDKQ-----NEVDSTSRALPPPLRFGKVDITLAS 751  
QY 290 -----PYVQCDSPANVECDGEISPAPPVTBEGNEDEDIDIGLLDNGCPANFEIDW 340  
Db 752 LAHDDLDLPAVPRIFSP-----PPLPKTPS-----CEFGDN-----EF 785  
QY 341 LLP-HGNCRDKYQCVHGNLVERRCGAGTHFSFELQOCDHIELVGCCTLPGGSESEVDDE 399  
Db 786 MFPKKSNNR-----VRGH--QSRPSTGS-----QLRNVPVPSIVTSGGR----- 822  
QY 400 DACTGWYCTEPIEW-----EPLNGCPADF--SIDHLLPHESDCGOYL----- 442  
Db 822 -----PALDEMASPSSIGHPLPSPADFNLSLWDFYEPH-----SYLSPAPEP 868  
QY 442 -----QCVHGQTIARPCGNLHF-----SPATQSCSPVTAGCOVFEC 480  
Db 869 QPSYEEESFNATVIHAPTPTSTATFGHPTISNVATPPLKQDVTESKASPADASATHOSS 928  
QY 481 S--DNQCTSTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 531  
Db 929 TGLTQEIITOLGNSNMLRPTKLRPSNDGRKASGPRPAAPSIPP-----PLPVSNISSPTS 984  
QY 532 AIPTPAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 585  
Db 985 EPPKDHPPSAPLSKVPVSTSPAAP-LARVPPVKLSSKAPPVPLPSADAPPVPPSTAPPV 1043  
QY 586 AAPTAAAP-----TAAPTA-----APTAVPEIPTVTSPTTAAPT-----TAAP 624  
Db 1044 PIPSTPPVPKSSGAPSPVPPVAPSPSEISIPAPSAPPVPPAPSGIPVPPKPSVAAPP 1103  
QY 625 APNTTVVPPTAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAAAPTAA 684  
Db 1104 VPKPSVAVPPVAPSGAPPVPKPSVAAPVPPSGAPPVPKPSVAAPVPPSGAPP--V 1161  
QY 685 HAPNTTAAPVTTTSA-----PATTPEDDDIDPPLN-----DPINP 720  
Db 1162 PKPSVAAPPVPAPSSGIPVPPKPAAGVPPVPPPSSE--APPVPKPSVGVPPVPP 1212

## RESULT 15

O13028 PRELIMINARY: PRT: 507 AA.  
AC O13028;  
DT 01-JUL-1997 (TREMBLREL. 04, CREATED)  
DT 01-JUL-1997 (TREMBLREL. 04, LAST SEQUENCE UPDATE)  
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)  
DE ANTIFREEZE GLYCOPROTEIN AFGP POLYPROTEIN PRECURSOR.  
OS BOREOGADUS SAIDA.  
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; ACTINOPTERYGII; NEOPTERYGII;  
OC TELEOSTEI; EUTELEOSTEI; PARACANTHOPTERYGII; GADIFORMES; GADOIDEI;  
OC GADIDAE; BOREOGADUS.  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE; 97286653.  
RA CHEN L., DEVRIES A.L., CHENG C.H.C.;  
RT "Convergent evolution of antifreeze glycoproteins in Antarctic  
notothenioid fish and Arctic cod";  
RL PROC. NATL. ACAD. SCI. U.S.A. 94:3817-3822(1997).  
DR EMBL; U43200; G2078483;  
KW SIGNAL; POLYPROTEIN.  
FT SIGNAL 1 33 POTENTIAL.  
FT CHAIN 34 507 ANTIFREEZE GLYCOPOLYPTIDE AFGP POLYPROTEIN.  
SQ SEQUENCE 507 AA; 44684 MW; 675D6FCB CRC32;

Query Match 9.2%; Score 405; DB 12; Length 507;  
Best Local Similarity 25.5%; Pred. No. 6.8e-06;  
Matches 172; Conservative 17; Mismatches 256; Indels 230; Gaps 22;

QY 87 LAGCTLPGP-----PAETTQA--PATTQAPTPTTQAPTPTTQAPTPTTQ 127  
Db 2 LTGCLLGPSCSTSPRYEQHOLLVARPAAARAATAATPATAATPAT---ATAATE 58



Search completed: July 18, 1999, 02:42:07  
Job time: 3630 sec

GenCore version 4.5  
Copyright (c) 1993 - 1998 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 18, 1999, 01:41:37 ; Search time 32.5 seconds  
(without alignments)  
1337.641 Million cell updates/sec

Title: US-09-294-663-3  
Perfect score: 4328  
Sequence: 1 MKTLLFLTALGLVAARPEV.....VFIPWKLDEIRQALNFEL 788

Scoring table: BLOSUM62

Searched: 180763 seqs, 55169189 residues

Database : SPTREMBL\_8.\*  
1: sp\_fungi.\*  
2: sp\_human.\*  
3: sp\_invertebrate.\*  
4: sp\_mammal.\*  
5: sp\_mhc.\*  
6: sp\_organelle.\*  
7: sp\_phage.\*  
8: sp\_plant.\*  
9: sp\_bacteria.\*  
10: sp\_rodent.\*  
11: sp\_virus.\*  
12: sp\_vertebrate.\*  
13: sp\_unclassified.\*  
14: sp\_archaea.\*

pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4328	100.0	788	018510	018510 trichoplusi
2	4298.5	99.3	807	018511	018511 trichoplusi
3	520.5	12.0	1404	092954	092954 homo sapien
4	461.5	10.7	3570	092952	092952 homo sapien
5	461.5	10.7	851	017893	017893 caenorhabdi
6	457	10.6	867	039782	039782 equine herp
7	452.5	10.5	866	039781	039781 equine herp
8	448.5	10.4	770	020908	020908 caenorhabdi
9	447.5	10.3	801	023635	023635 caenorhabdi
10	442.5	10.2	373	076810	076810 anopheles g
11	414.5	9.6	2378	091365	091365 caenorhabdi
12	408	9.4	825	017921	017921 caenorhabdi
13	393	9.1	1777	022579	022579 caenorhabdi
14	392	9.1	1188	041805	041805 zea mays (m
15	388	9.0	750	039307	039307 equine herp
16	387.5	9.0	622	014881	014881 homo sapien
17	383	8.8	507	013028	013028 boreogadus
18	371	8.6	1611	042854	042854 schizosacch
19	367	8.5	1151	057580	057580 gallus gall
20	366	8.5	1131	018529	018529 caenorhabdi
21	362	8.4	412	077242	077242 heterodera
22	361.5	8.4	1118	028226	028226 canis famil
23	351.5	8.1	2187	070670	070670 mus musculus
24	351	8.1	846	001699	001699 caenorhabdi
25	345	8.0	232	027423	027423 drosophila
26	340.5	7.9	353	025677	025677 podocoryne
27	340.5	7.9	584	017802	017802 caenorhabdi
28	340	7.9	279	014888	014888 homo sapien
29	339.5	7.8	901	044562	044562 actinomyces

30	328	7.6	379	3	027929	027929 drosophila
31	327	7.6	2225	3	045881	045881 caenorhabdi
32	327	7.6	2761	3	019522	019522 caenorhabdi
33	326	7.5	5376	10	088799	088799 mus musculu
34	324	7.5	1229	3	094185	094185 caenorhabdi
35	324	7.5	528	4	029071	029071 sus scrofa
36	323	7.5	928	12	098906	098906 gallus gall
37	323	7.5	722	12	013083	013083 dissostichu
38	322	7.4	216	4	028501	028501 macaca mula
39	321.5	7.4	350	8	041719	041719 zea diplope
40	314.5	7.3	303	8	041814	041814 zea mays (m
41	314.5	7.3	3247	11	065553	065553 bovine herp
42	314	7.3	328	8	042366	042366 zea mays (m
43	313	7.2	946	8	022015	022015 cylindrothe
44	313	7.2	907	11	066537	066537 epstein-bar
45	310	7.2	886	8	022016	022016 cylindrothe

ALIGNMENTS

RESULT 1						
018510						
ID	018510	PRELIMINARY;				
AC	018510;		PRT;	788 AA.		
DT	01-JAN-1998	(TREMBLREL. 05, CREATED)				
DT	01-JAN-1998	(TREMBLREL. 05, LAST SEQUENCE UPDATE)				
DE	01-NOV-1998	(TREMBLREL. 08, LAST ANNOTATION UPDATE)				
OS	INSECT INTESTINAL MUCIN IIM14.					
OS	TRICHOPLUSTIA NI (CABBAGE LOOPER).					
OC	EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;					
OC	PTERYGOTA; LEPIDOPTERA; NOCTUIDAE; NOCTUIDAE; TRICHOPLUSTIA.					
RP	{1}					
RP	SEQUENCE FROM N.A.					
RX	WANG P.; GRANADOS R.R.;					
RA	"Molecular cloning and sequencing of a novel invertebrate intestinal					
RT	mucin cDNA."					
RL	J. BIOL. CHEM. 272:16663-16669(1997).					
DR	EMBL; AF000605; G224919;					
SQ	SEQUENCE 788 AA; 81716 MW; 8F8ADE83 CRC32;					
Query Match	100.0%;	Score 4328;	DB 3;	Length 788;		
Best Local Similarity	100.0%;	Pred. No. 4.2e-120;				
Matches 788;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;		
QY 1	MKTLFLTALGLVAARPEVSDAENKPNALHPEHPDGP	PAEQHLLPHEYDCTKFFYCEYG	60			
DB 1	MKTLFLTALGLVAARPEVSDAENKPNALHPEHPDGP	PAEQHLLPHEYDCTKFFYCEYG	60			
QY 61	LKFIAPRDCAPGTEFKFSQAQCVHAALAGCTLP	GPAAETTPAATTPAATTPAATTTTQ	120			
DB 61	LKFIAPRDCAPGTEFKFSQAQCVHAALAGCTLP	GPAAETTPAATTPAATTPAATTTTQ	120			
QY 121	APTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ	APTTTQAPTTTQAPTTTQAPTTTQ	180			
DB 121	APTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ	APTTTQAPTTTQAPTTTQAPTTTQ	180			
QY 181	PTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ	PTTQAPTTTQAPTTTQAPTTTQAPTTTQ	240			
DB 181	PTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ	PTTQAPTTTQAPTTTQAPTTTQAPTTTQ	240			
QY 241	PICELLPNCPDAFDHLLPHDKYCNLFYQCSNGY	FEQRCPEGLYFNFYVORCDSPAN	300			
DB 241	PICELLPNCPDAFDHLLPHDKYCNLFYQCSNGY	FEQRCPEGLYFNFYVORCDSPAN	300			
QY 301	VECDGEISAPVTEGNEDEDIDGLDNGCPANFEID	WLLPHGNRCCKYQCCHVGNLV	360			
DB 301	VECDGEISAPVTEGNEDEDIDGLDNGCPANFEID	WLLPHGNRCCKYQCCHVGNLV	360			
QY 361	ERRCGAGTHFSFELOQCCHIELVGTCTLPGSEEV	VDVEDACTGWYCTPEIWEPLNG	420			

```
Db 361 ERRCAGTHSFELQOCDHIELVGCCTLPGGSESEVDEDACTGWCYPTPIEWELPNG 420
QY 421 CPADFSIDHLLPHESDCGQYLCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480
Db 421 CPADFSIDHLLPHESDCGQYLCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480
QY 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540
Db 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540
QY 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600
Db 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600
QY 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 660
Db 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 660
QY 661 PPTVAHAPNTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 720
Db 661 PPTVAHAPNTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 720
QY 721 VCDGNNVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSSEYGVQVFI 780
Db 721 VCDGNNVLVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMSSEYGVQVFI 780
QY 781 ROALNFEL 788
Db 781 ROALNFEL 788

RESULT 2
O18511 ID O18511 PRELIMINARY; PRT; 807 AA.
AC O18511;
DT 01-JAN-1998 (TREMBREL. 05, CREATED)
DT 01-JAN-1998 (TREMBREL. 05, LAST SEQUENCE UPDATE)
DE 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE INSECT INTESTINAL MUCIN IIM22.
OS TRICHOPLUSIA NI (CABBAGE LOOPER).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; LEPIDOPTERA; NOCTUIDAE; NOCTUIDAE; TRICHOPLUSIA.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE; 97341213.
RA WANG P.; GRANADOS R.R.;
RT "Molecular cloning and sequencing of a novel invertebrate intestinal
RT mucin cDNA.";
RL J. BIOL. CHEM. 272:16663-16669(1997).
DR EMBL; AF00606; G224921; -.
SO SEQUENCE 807 AA; 83408 MW; 15E08C34 CRC32;

Query Match 99.3%; Score 4298.5; DB 3; Length 807;
Best Local Similarity 97.5%; Pred. No. 3.1e-119;
Matches 787; Conservative 0; Mismatches 1; Indels 19; Gaps 1;

QY 1 MIKTLLELTLGLVAARPEVSDAEKNALHPHDPDCPPAEQHWLLPHEYDCTKEYCEYG 60
Db 1 MIKTLLELTLGLVAARPEVSDAEKNALHPHDPDCPPAEQHWLLPHEYDCTKEYCEYG 60
QY 61 LKFIAPRDCAPGTEFKESAQTCVHAALAGCTLPGPRAETTQAPATTQAPTTTQ 120
Db 61 LKFIAPRDCAPGTEFKESAQTCVHAALAGCTLPGPRAETTQAPATTQAPTTTQ 120
QY 121 APTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180
Db 121 APTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 180
QY 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 240
Db 181 PTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQ 240
```

```
QY 241 PICELLPNGCPADFDIHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNPYVQCDSPAN 300
Db 241 PICELLPNGCPADFDIHLIPHDKYCNLFYQCSNGYTFEQRCEGLYFNPYVQCDSPAN 300
QY 301 VECDEGISAPPVTEGNEDEDIDIGLLONGCPANFEIDWLLPHGNRCCKYYOCVHGNLV 360
Db 301 VECDEGISAPPVTEGNEDEDIDIGLLONGCPANFEIDWLLPHGNRCCKYYOCVHGNLV 360
QY 361 ERRCGAGTHSFELQOCDHIELVGCCTLPGGSESEVDEDACTGWCYPTPIEWELPNG 420
Db 361 ERRCGAGTHSFELQOCDHIELVGCCTLPGGSESEVDEDACTGWCYPTPIEWELPNG 420
QY 421 CPADFSIDHLLPHESDCGQYLCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480
Db 421 CPADFSIDHLLPHESDCGQYLCVHGQTIARPCPGNLHFSPTQSCSPVTAGCQVFEC 480
QY 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540
Db 481 SDNOCSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 540
QY 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600
Db 541 APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 600
QY 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 641
Db 601 TTAVPEIPTVTSPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 660
QY 642 APAPNTVTVPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 701
Db 642 APAPNTVTVPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 701
QY 702 CVEECNVLPWAHADCKYWCDCNNQVNVVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761
Db 702 CVEECNVLPWAHADCKYWCDCNNQVNVVVCSEGLQFNPTTKTCDFACNVGCVRSNIQMS 761
QY 762 ESEYGVQVFIWPKNKLDEDIRQALNFEL 788
Db 762 ESEYGVQVFIWPKNKLDEDIRQALNFEL 807

RESULT 3
O92954 ID O92954 PRELIMINARY; PRT; 1404 AA.
AC O92954;
DT 01-FEB-1997 (TREMBREL. 02, CREATED)
DT 01-FEB-1997 (TREMBREL. 02, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBREL. 08, LAST ANNOTATION UPDATE)
DE MEGAKARYOCYTE STIMULATING FACTOR.
OS HOMO SAPIENS (HUMAN).
OC EUKARYOTA; METAZOA; CHORDATA; VERTEBRATA; MAMMALIA; EUTHERIA; PRIMATES;
OC CATARRHINI; HOMINIDAE; HOMO.
RN [1]
RP SEQUENCE FROM N.A.
RA TURNER K.J.; FITZ L.J.; TEMPLE P.; JACOBS K.; LARSON D.; LEARY A.C.;
RA KELLEHER K.; GIANNOTTI J.; CALVETTI J.; FITZGERALD M.; KRIZ M.J.;
RA FERENZ C.; GROBHOZ J.; FRASER H.; BEAN K.; NORTON C.R.; GESNER T.;
RA BHATIA S.; KRIZ R.; HEWICK R.; CLARK S.C.;
RL BLOOD 78:279-279(1991).
RN [2]
RP SEQUENCE FROM N.A.
RA MERBERG D.M.; FITZ L.J.; TEMPLE P.; GIANNOTTI J.; MURTHA P.;
RA FITZGERALD M.; SCALTRETO J.; KELLEHER K.; PREISSNER K.; KRIZ R.;
RA JACOBS K.; TURNER K.;
RL (IN) PREISSNER K.T.; ROSENBLATT S.; KOST C.; WEGERHOFF J.;
RL MOSHER D.F. (EDS.); BIOLOGY OF VITRONECTINS AND THEIR RECEPTORS.;45-52;
RL ELSEVIER SCIENCE PUBLISHERS B.V. (1993).
RN [3]
RP SEQUENCE FROM N.A.
RA TURNER K.J.; FITZ L.J.; TEMPLE P.; JACOBS K.; LARSON D.; LEARY A.C.;
RA KELLEHER K.; GIANNOTTI J.; CALVETTI J.; FITZGERALD M.; KRIZ M.J.;
RA FERENZ C.; GROBHOZ J.; FRASER H.; BEAN K.; NORTON C.R.; GESNER T.;
RA BHATIA S.; KRIZ R.; HEWICK R.; CLARK S.C.;
```





[illegible]

```

Db 427 SVSWITVWLTSLIGLPGFNNNNFDIIINLANCRVPIGEFHIGYISTPCNNOITIIIVS 486
QY 445 -----HQQTIAKPCGNLHSPATQSCSPV 470
Db 487 GIELIGLTFNFMFMIGGLIYTWKTLTIFRFLSNFKVMKWTSLHLLLT----- 540
QY 471 TACQVFECDSNQCSTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 530
Db 540 -----YKLDVINSIAT-SPTADPTTTTTEATTTTTEATTTTTEATTTTTEATTTT 592
QY 531 TAIPTPAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTA 590
Db 593 TSTSTTTTTTTTTTTT-----PPTTSTSTTTTTTTTTTTTTTTTTTTTTTTTTTAPT 648
QY 591 APTAAPTAAPTAV-----PEIPTVTSP---PTAAPTAAAPNTVTVPTAAAPT 639
Db 649 TTTTTPCNPNAVSLMGDANNQIFIDVYNNYLSLTDPTVGTIVSTMTVTCALNGYN 708
QY 640 T-----AAPNTVTVPTAAAPT--AAPTVAHAPNTTAAPVTTSA 680
Db 709 TYMLFNGGGGADNQ--NLPTQISITLVCTSDTMVMNVEVINGVYTRA 757

RESULT 9
Q23635 PRELIMINARY: PRT: 801 AA.
ID Q23635;
AC Q23635;
DT 01-NOV-1996 (TREMREL. 01, CREATED)
DT 01-NOV-1996 (TREMREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE SIMILAR TO LONG TANDEM REPEAT REGION OF STALIDASE.
GN ZK84.1.
OS CAENORHABDITIS ELEGANS.
OC EUKARYOTA; METAZOA; NEMATODA; SECERNENTEA; RHABDITIA; RHABDITIDA;
OC RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-BRISTOL N2;
RX MEDLINE; 94150718.
RA BURTON J., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,
RA WILSON R., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,
RA DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,
RA HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,
RA KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,
RA MCMURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,
RA RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,
RA SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MIEG J., THOMAS K.,
RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RL elegans.";
RL NATURE 368:32-38(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA KIRSTEN J.;
RL SUBMITTED (APR-1995) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U23181; G726436; -.
SQ SEQUENCE 801 AA; 77123 MW; 7288A89D CRC32;

Query Match 10.3%; Score 447.5; DB 3; Length 801;
Best Local Similarity 26.4%; Pred. No. 3.8e-07;
Matches 193; Conservative 41; Mismatches 288; Indels 209; Gaps 28;

QY 10 ALGLVAARPE--VSDA-EKNPALHEPHDCPPAQQHLLPHEVDCIKFYCYEGLKFIAP 66
Db 194 ASGVTVSEQGAGADPEAPVAVVEETPAPTAAEE-----TPAPATEASNAVGP 243

QY 67 RDCAPGTEFEKFSQTCVHAALAGCTLPGPAAETTOAPTATTOAPT--TTOAPT 125
Db 244 EGYVDGTA---NAASVAEAPVETPAPAPAEETPAPTATSEAPAPAPTAEPTAPETV 300

```

```

QY 126 TOAP-----TTTTQAPTQAPTQAPTQAPTQAPTQAPTQAPTQAPTQAPTQAPT 173
Db 301 SAAPEAANSYDSAGGDAATAPAPSSSEADAAAPTDSAAADTAALVDTSSSEHAESTE 360
QY 174 APTTTQAPT--TTOAPT--QAPTITQAAATPAA--TTPAATTTTAAATTPAAT 221
Db 361 APATDIAATETTPASVPAPVADAAGYDPSPIEETPAPAAEDTPAPASAAEETPAP 420
QY 222 TPAATTPGVPAPASVPVPPICELLPNGCPADF---DIHLLPHDKYCNLFYQCSNGYT 277
Db 421 APAEE--TPAETASAP--DAAGRAPADVAAPADVATTAPE-----T 461
QY 278 FEQRCEGLYFNPYQVRCDSANVECDGSIAPPV---TEGNEDEDIDIGLLDNGCPA 334
Db 462 SSAQSAAGSYDVP-----SEPAS-----EVT-APIVESATEAPSDSAAPIGPAASE 510
QY 335 NFEIDWLLPHGNRCDKYQCVHGNLVERCGAGTHSFELQOCDHIELVGCITLPGGESEE 394
Db 511 -----PIEAPATDAATLETAPAPAAEPAP 534
QY 395 VDVEDACTGWYCPTETPIEWEPLPNCPADFSIDHLLPHESDCGYLCVHGQTIARPCP 454
Db 535 A---AEAAGYDAPSSVPEETAP----- 556
QY 455 GNLHSPATQSCSPVTCQVFECDSDNQCSTSTAAPTAAPTAAPTAAPTAAPTAAPT 514
Db 556 -----APAADETAP-----APAAEETPAPAPAAEETPAPAPAAE 590
QY 515 VPPATPPA---TAAPVPPPTTAIPTAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 571
Db 591 ETAPATAAAEETPAPAPAADETAPAPAAEETPAPAPAVEETPAPA---PAVEETPAPAP 647
QY 572 AVPEIPTVTSAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 631
Db 648 AAEEPT--APAPAAEYAAPVAEETPAPAPAAEETP-----APAAEETPAPAPAAEET 700
QY 632 VPPTAAPTAAAPNTVTVPTTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPT 675
Db 701 -PASAPAAEETPAPAPAAEAPAPAP-AAETPAPAAEGAAAPATVSSGYDAAQSTDV 758
QY 676 TTTSAPATPE 686
Db 759 VASSAPATSSD 769

RESULT 10
O76810 PRELIMINARY: PRT: 373 AA.
ID O76810;
AC O76810;
DT 01-NOV-1998 (TREMREL. 08, CREATED)
DT 01-NOV-1998 (TREMREL. 08, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMREL. 08, LAST ANNOTATION UPDATE)
DE ICHIT PROTEIN.
GN ICHIT.
OS ANOPHELES GAMBIAE (AFRICAN MALARIA MOSQUITO).
OC EUKARYOTA; METAZOA; ARTHROPODA; TRACHEATA; HEXAPODA; INSECTA;
OC PTERYGOTA; DIPTERA; NEMATOCERA; CULICOIDEA; CULICIDAE; ANOPHELES.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SUAKOKO; TISSUE-MAINLY MIDGUT;
RA DIMOPOULOS G., SEELEY D., WOLF A., KAFATOS F.C.;
RT "Malaria infection of the mosquito Anopheles gambiae activates immune
RT responsive genes during critical transition stages of the parasite
RT life cycle.";
RL SUBMITTED (SEP-1998) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; AJ010903; E1320052; -.
SQ SEQUENCE 373 AA; 40647 MW; A7231CDB CRC32;

Query Match 10.2%; Score 442.5; DB 3; Length 373;
Best Local Similarity 36.1%; Pred. No. 3.1e-07;
Matches 137; Conservative 23; Mismatches 124; Indels 95; Gaps 20;

```





```

RA VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,
RA WILKINSON-SPROAT J., WOHLDMAN P.;
RT *2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans.";
RT NATURE 368:32-38(1994).
RN [2]
RN
RN
RN SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RA NHAN M., HAWKINS J.;
RL SUBMITTED (APR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
RN [3]
RN
RN SEQUENCE FROM N.A.
RC STRAIN=BRISTOL N2;
RC STRAIN=BRISTOL N2;
RA WATERSTON R.;
RL SUBMITTED (MAR-1996) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; U51598; G1255319;
SQ SEQUENCE 825 AA; 86617 MW; 90FA8FCE CRC32;

```

```

Query Match          9.4%; Score 408; DB 3; Length 825;
Best Local Similarity 24.0%; Pred. No. 5.4e-06;
Matches 148; Conservative 26; Mismatches 159; Indels 284; Gaps 17;

QY 102 APATTAQATTTQATTTTQATTTTQATTTTQATTTTQATTTTQATTTTQATTTTQATTTTQAP 161
   || || || || || || || || || || || || || || || || || || || || ||
Db 460 APPTTCGTEYPVTVTSTMSPPPTTVTPPTPTPVPTTNTP---PANPTTATPTTGTGSK 516

QY 162 TTTQAP-----TTTQATTTTQATTTTQATTTTQATTTTQATTTTQATTTTQATTTTAA 210
   || || || || || || || || || || || || || || || || || || || || ||
Db 517 OTNTISHLSTIGSVITSTPTMAPQTSASPTPTTHTTASQPTT----TKPVVTTNSV 571

QY 211 TTPAATTAATTAATTAATPGVP-APTSAPVMPPICELLPNCPCPADFDIHLIPHDKYCNLF 269
   || || || || || || || || || || || || || || || || || || || || ||
Db 572 TPTGHT--TVPVPTTGGSTTQTATPVKP----- 601

QY 270 YQCSNGYTFEQRCEGLYFNYPVQRCDSANVECDGEISPAPPVTEGNEDEDIDIGDLLD 329
   || || || || || || || || || || || || || || || || || || || || ||
Db 601 ----- 601

QY 330 NGCPANFEIDWLLPHGNRCDKYYQCVHGNLVERCGAGTHSFELQOCDHIYGCILPG 389
   || || || || || || || || || || || || || || || || || || || || ||
Db 601 -----TVP- 604

QY 390 GESEEVDDVEDACTGWYCPTETIEWEPLNCPADFSIDHLLPHESDCGQYLQCVHGQTI 449
   || || || || || || || || || || || || || || || || || || || || ||
Db 604 ----- 604

QY 450 ARPCPNLHFSPATQCESPVTAGCVFECSDNQCTSTAAPTAAPTAAPTAAPTAA 509
   || || || || || || || || || || || || || || || || || || || || ||
Db 604 -----SSTTQTAPPVTVTPTSQPPVTTTS 626

QY 510 APSTVVPPATPPATAAPVPTTAIPTAPTAAPTAAPTTA-APESPTTVTVPPTAAPTAA 568
   || || || || || || || || || || || || || || || || || || || || ||
Db 627 LLTTLTP--TVPVTTTVVPSATVPPTPTTVTVAATTTSKAP---VVTTSPTLAPT-S 680

QY 569 PTTAVPEIPTVTSAPTA-APTAAPTAAP-----TAAPTAVPEIPTVTSPTTAAAP-T 620
   || || || || || || || || || || || || || || || || || || || || ||
Db 681 PTKLPTSPSTVGTSGTPAPANLTPTTAPVNPSTSTAPT--APVNPSTPTATVPVPT 738

QY 621 TAAPAPNTTVTVPTTAAPTAAAPAPNT-TVTVPTTAAPTAAAPT-VAHAPNTTAAPVTT 678
   || || || || || || || || || || || || || || || || || || || || ||
Db 739 TTTPTTTTSTTTTTTTTTTTTTTTTQTTPTTPVTTPTSTITPTTRPVTPASQAPATPTTPT 798

QY 679 SAPATPDEDDDDIDPLP 695
   || || || || || || || || || || || || || || || || || || || || ||
Db 799 TIPTTQFSSAVSVALP 815

```

```

RESULT 13
Q22579
ID Q22579 PRELIMINARY; PRT; 1777 AA.
AC Q22579;
DT 01-NOV-1996 (TREMBLREL. 01, CREATED)

```

01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)	Query Match	9.1%;	Score 393;	DB 3;	Length 1777;
01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)	Best Local Similarity	23.8%;	Pred. No. 2.5e+05;		
01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)	Matches 205;	Conservative 92;	Mismatches 283;	Indels 282;	Gaps 36;
SIMILAR TO A. FAECALIS POLY(3-HYDROXYBUTYRATE) DEPOLYMERASE.					
T19D12.1.					
CAENORHABDITIS ELEGANS.					
EUKARYOTA; METAZOA; NEMATODA; SECCERNENTER; RHABDITIA; RHABDITIDA;					
RHABDITINA; RHABDITOIDEA; RHABDITIDAE; PELODERINAE; CAENORHABDITIS.					
[1]					
SEQUENCE FROM N.A.					
MEDLINE: 94150718.					
WILSON R., AINSCOUGH R., ANDERSON K., BAYNES C., BERKS M., BONFIELD J.,					
BURTON J., CONNELL M., COPSEY T., COOPER J., COULSON A., CRAXTON M.,					
DEAR S., DU Z., DURBIN R., FAVELLO A., FULTON L., GARDNER A., GREEN P.,					
HAWKINS T., HILLIER L., JIER M., JOHNSTON L., JONES M., KERSHAW J.,					
KIRSTEN J., LAISTER N., LATREILLE P., LIGHTNING J., LLOYD C.,					
MCURRAY A., MORTIMORE B., O'CALLAGHAN M., PARSONS J., PERCY C.,					
RIFKEN L., ROOPRA A., SAUNDERS D., SHOWNKEEN R., SWALDON N., SMITH A.,					
SONNHAMMER E., STADEN R., SULSTON J., THIERRY-MTEG J., THOMAS K.,					
VAUDIN M., VAUGHAN K., WATERSTON R., WATSON A., WEINSTOCK L.,					
WILKINSON-SPROUT J., WOHLDMAN P.;					
"2.2 mb of contiguous nucleotide sequence from chromosome III of C.					
elegans";					
NATURE 368:32-38(1994).					
[2]					
SEQUENCE FROM N.A.					
FAVELLO T.;					
SUBMITTED (DEC-1995) TO EMBL/GENBANK/DBJ DATA BANKS.					
[3]					
SEQUENCE FROM N.A.					
WATERSTON R.;					
SUBMITTED (NOV-1995) TO EMBL/GENBANK/DBJ DATA BANKS.					
EMBL: U41263; G1086828; -					
SEQUENCE 1777 AA; 179736 MW; 24064E32 CRC32;					
QY	52	KFIAPDCAPGTEFKESAQTCVHAAGCTLPGPP-----	-----AETQAPAT 105		
DB	187	KTIIOGDC-PST---VAPVTC-----	STMPMPTOPLVTSTAPEVTTVAQTTPAIV 235		
QY	106	TOAPTITQAPTTT-----	TOAPTITQAPTTTQAPTTTQAPTT----- 145		
DB	236	TTANTITQGVTTAGTTTTVTRAQNSLAAATTAPSTNTTQGVTTTGVKTTTVAQNS 295			
QY	145	TOAPTITQAPTTTQAPTTTQA-----	PTTITQAPTTTQAPTTTQAPT 194		
DB	296	TWAATTTASNTTTPQVTTTSTQGISGTTTAQATPSSSVIPTTTQ---TTQRPSTGIPS 352			
QY	195	TI-----	TOAATTPAAT-----TPAATTPAATPAA----- 221		
DB	353	TVSTSGTSGTSGIPSTTQTSSAPSTYTSNFTPSPTTLLTSTIAPSTQGVPTSSKSS 412			
QY	221	--TTPAAT--TPGVP-----	APTSAPVPPICELLNGCPADEFILLIHPDK- 265		
DB	413	PNSTPTTITPGAPSTLSSSSSTIVSTIITPTPKVSLTLSQSPPTSTPLVWSSSS 472			
QY	265	-----YCNLFYQCSNGY-TFEQRCPEGFLYFVQRCD-----	----- 297		
DB	473	GSSTVTVTSTIIPSTQGVPTSTSNQPTPSNSTPKSTVTASPTTGTATSTASPSITIS 532			
QY	297	-----	SPANVECDGETSPAPVTEGNEDED 321		
DB	533	SAPTSQSHSPSSTMTSTVPVTSTFTASTTTTVPITVAPGQCYSQSNVAVAFETSGTSDLD 592			
QY	322	IDI-----	GDLNDC-----PANFEIDWLLPHGNK----- 348		
DB	593	LDIQNFIANVLFYSCAPVILGDLTNRFTAISLVPPFNDTLDLMTYGAETPSGISAA 652			
QY	348	CDKYQCYVGHNLVYRRCGAGTHFSFELQQCDHIELVG-----	CTLPGGESVEVDDEACT 403		

```

Db 653 LDTFNILARGNAV-----ISDAFNIPNITRKGQGFVILVANSDESQASVDSAT 703
QY 404 -----GWYCPTETIEWELPNCGPADFSIDLPHESDCGOVLCQVHGQTIARPCPGNLH 458
Db 704 NLKAQGFNITVAFKSGGFDVLASQPSYNTIVQDAD-KRYVATLIGNVLTNYCNLA 762
QY 459 FSPATOSCSPVTTAGOVCECDNCTSTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 518
Db 763 ASSTT-----MTSGVYSVTVTTCGSSQAPSSVT-VPTGTGTS--AASTGSGITS 813
QY 519 TPATAAPVPPTTAIPTAAPTAA-----TAAPTAAPESTPTVTPPTAAPTAAPTAA 575
Db 814 TQOATS-----TSSVITGTSAPQSTAVSSITTSSTTGTGTPAQSTVASITTVS- 868
QY 576 IPTVTS-----APTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAAPTAA 617
Db 868 -PYTTECICITVNSNFTGTSTOGLTSSAQSTASTGVSTVASSTIPO-GSSSSSP--Q 923
QY 618 APTTAAPANT-----TVTPPTAAPT-----TAAPANTTVTPPTAAPTAAPTAA 666
Db 924 SPTSAPQAOSSTSAATVSSQSPTSPQAQOSTPAGSSTVTVOQSSFSQSQSTTQI 983
QY 667 APNTTA-----APVTTTSAPATT 684
Db 984 GSSTTVPTQAPSSSTSGGTTT 1005

RESULT 14
Q41805 PRELIMINARY; PRT: 1188 AA.
AC Q41805;
DT 01-JAN-1996 (TREMBLREL. 01, CREATED)
DT 01-NOV-1996 (TREMBLREL. 01, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE EXTENSIN-LIKE PROTEIN PRECURSOR.
OS ZEA MAYS (MAIZE)
OC EUKARYOTA; VIRIDIPLANTAE; CHAROPHYTA/EMBRYOPHYTA GROUP; EMBRYOPHYTA;
OC TRACHEOPHYTA; EUPHYLOPHYTES; SPERMATOPHYTA; MAGNOLIOPHYTA; LILIOPSIDA;
OC POALES; POACEAE; ZEA.
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=B73; TISSUE-POLLEN;
RA RUBINSTEIN A.L., BROADWATER A.H., LOWREY K., BEDINGER P.A.;
RL SUBMITTED (JUN-1994) TO EMBL/GENBANK/DBJ DATA BANKS.
DR EMBL; Z34465; G600118;
DR PFM; PF00560; LRR; 3.
KW SIGNAL.
FT SIGNAL.
SQ SEQUENCE 1188 AA; 120980 MW; 47E3D9BA CRC32;

Query Match 9.1%; Score 392; DB 8; Length 1188;
Best Local Similarity 22.5%; Pred. No. 2e-05;
Matches 170; Conservative 87; Mismatches 324; Indels 176; Gaps 28;

QY 2 IKTLFETALGLVAARPEVSDAEKNALHEPHDC-----PRAEQHLLPHEVDCTKFFY 56
Db 520 VKITSPAPIGSPSPVPVSVSPVPVKSPPPPAPVGSPPPEKSPPPP-----570
QY 57 CEYGLKFIAPRCACPTKFSQATCVHAALAGCTLPGPPAETQAPATQAP-----110
Db 570 -----APVASPPPPVKSPPPTLVASPPPPVKSPPPPAPVASPPPPVKSPPPTVA 621
QY 110 -TTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 168
Db 622 SPPPPAPVASSPPPMKSPPPPTVSSPPPEKSPPPPPPAKSTPPPEYPTPTTSVKSSP 681
QY 169 TTTQAPTTTQAP-----TTTQAPTTTQAPTTTQAPTTTQAPTTTQAPTTTQAP 215
Db 682 PPKSLPPTLLPSPPQKPTPPPTSPKSPSPKSPKSPKSPKSPKSPKSPKSPKSPKSP 741
QY 215 -----ATTPAATTAAATPGV-----PAPTSAPVWPPICELLNPGCPADFDHLLPH 262

```

```

Db 742 SSPPPTPVSSPPALAPVSSPPSVKSSPPPPAPLSSP--PPAPQV--KSSPPPVQVSSPPPA 797
QY 263 DKYCNLFYQCSNGYTFEQRC-PEGLYFNFYQVQCDSPANVECDGEISPAAPPVTEGNEDED 321
Db 798 PKSSPPLAPVSPQVQEKTSPPAPLSSPPLAPKSSPPH---VVSSPPPVVKSPP--852
QY 322 IDIGDLLNGCCPANTEIDWLLPHGNRCDKYOCVGNLVERRCGAGTHFSFELQOCDHIE 381
Db 852 -----PAPVSSPPLTP-----KPASPPPAHVS-----SPPE 876
QY 382 LVGCTLPGESEEDVDEDACTGWCTEPIEWEPLNPGCPADFSIDLPHESDCGOYL 441
Db 877 VKPSTPPAPTIVISP-----PSEPKSSPP-----PTVSLPPLPVIVKSS-----916
QY 442 QCVHGQTIARPCGNLHFSPT-QSCESPVTAGCQVFECDSDNQCSTSTAAPTAAPTAA 500
Db 916 -----PPAMVSSPMTKSSPPVVV-----SSPPPTVKSSPPAPVSSPP 957
QY 501 AAPTAAPTAAPTVTPP-----ATPPATAAPVPTTAIPTAAPTAAPTAAPTAAPTAA 557
Db 958 ATPKSSPPAPVNLPPPEVKSSPPPTPVSSPP-----PAPKSSPPAPAPMSSPP--PPEV 1009
QY 558 TVPPTAAPTAAPTAV--PEIPITVTSAB--TAAPTAAPTAA-----TAAPTAVPEI 607
Db 1010 KSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPVKSPPPPAPVSSPPPPAPVSSPP 1069
QY 608 PTTVTSPTAAPTAAAPANTTVTPPTAAPTAAAPANTTVTPPTAAPTAAPTAAPTAA 667
Db 1070 PPKVKSPPPPAPVSSPPPP--VKSPPPAPVSSPPPP--IKSPPPPAPVSSPPPPAPVK 1123
QY 668 PNTTAAPTVTTT-APATTP-----EDDDIDPLPNDP 698
Db 1124 PPSLPPAPVSSPPPPVTPVTPAPPKKEQSLPPAESOP 1160

RESULT 15
O39307 PRELIMINARY; PRT: 750 AA.
AC O39307;
DT 01-JAN-1998 (TREMBLREL. 05, CREATED)
DT 01-JAN-1998 (TREMBLREL. 05, LAST SEQUENCE UPDATE)
DT 01-NOV-1998 (TREMBLREL. 08, LAST ANNOTATION UPDATE)
DE POSITIONAL COUNTERPART OF HSV-1 GENE US5.
GN 71.
OS EQUINE HERPESVIRUS 4.
OC VIRUSES; DSDNA VIRUSES; NO RNA STAGE; HERPESVIRIDAE;
OC ALPHAHERPESVIRINAE; VARICELLOVIRUS.
[1]
RN SEQUENCE OF 685-750 FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE; 93389454.
RA CULLINANE A.A., NEILAN J., WILSON L., DAVIDSON A.J., ALLEN G.;
RT "The DNA sequence of the equine herpesvirus 4 gene encoding
RT glycoprotein gpi1/18, the homologue of herpes simplex virus
RT glycoprotein gD."
RL J. GEN. VIROL. 74:0-0(0).
RN [2]
RN SEQUENCE OF 1-111 FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE; 93119267.
RA NAGESHA H.S., CRABB B.S., STUDDERT M.J.;
RT "Analysis of the nucleotide sequence of five genes at the left end of
RT the unique short region of the equine herpesvirus 4 genome."
RL ARCH. VIROL. 128:143-154(1993).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=NS80567;
RX MEDLINE; 98264497.
RA TELFORD E.A.R., WATSON M.S., PERRY J., CULLINANE A.A., DAVIDSON A.J.;
RT "The DNA sequence of equine herpesvirus-4."
RL J. GEN. VIROL. 79:1197-1203(1998).
RN [4]
RN SEQUENCE FROM N.A.

```

